richardwu.ca

STAT 330 Course Notes

Mathematical Statistics

Yingli Qin • Fall 2018 • University of Waterloo

Last Revision: December 12, 2018

Table of Contents

Abstract

These notes are intended as a resource for myself; past, present, or future students of this course, and anyone interested in the material. The goal is to provide an end-to-end resource that covers all material discussed in the course displayed in an organized manner. These notes are my interpretation and transcription of the content covered in lectures. The instructor has not verified or confirmed the accuracy of these notes, and any discrepancies, misunderstandings, typos, etc. as these notes relate to course's content is not the responsibility of the instructor. If you spot any errors or would like to contribute, please contact me directly.

1 September 7, 2018

1.1 Random variables

We have two types (not include mixture r.v.s) random variables (r.v.s):

Discrete Probability (mass) function of X

$$f(x) = P(X = x)$$

Support set of X

$$A = \{x \mid f(x) > 0\}$$

The following two conditions must hold

- $$f(x) \geq 0$$
- $$\sum_{x \in A} f(x) = 1 \quad \text{or} \quad \sum_{x \in \mathbb{R}} f(x) = 1$$

Continuous Probability density function (pdf) of X

$$f(x) = \frac{d}{dx} F(x) = F'(x)$$

if F is differentiable at x, otherwise $f(x) = 0$.

Support set of X

$$A = \{x \mid f(x) > 0\}$$

The following two conditions must hold

- $$f(x) \geq 0 \quad \forall x \in \mathbb{R}$$
- $$\int_{x \in A} f(x)\,\mathrm{d}x = 1 \quad \text{or} \quad \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$$

Some examples of **discrete** r.v.s

Bernoulli $X \sim Bernoulli(p)$ for $0 < p < 1$ where

$$P[X = 1] = p \quad \text{or} \quad P[X = 0] = 1 - p$$

therefore

$$f(x) = P[X = x] = p^x(1-p)^{1-x} \quad x = 0, 1$$

and $A = \{0, 1\}$.

Binomial $X \sim BIN(n, p)$ for $n = 1, 2, \ldots$ and $0 < p < 1$. X represents the number of successes of n iid $BERN(p)$ trials or X (or X is sum of n iid $BERN(p)$ r.v.s):

$$X = \sum_{i=1}^{n} Y_i \quad Y_i \sim BERN(p)$$

therefore

$$f(x) = P[X = x] = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \ldots, n$$

and $A = \{1, 2, \ldots, n\}$.

Geometric $X \sim GEO(p)$ for $0 < p < 1$. X represents the number of failures before the 1st success in a sequence of iid $BERN(p)$ trials, therefore

$$f(x) = P[X = x] = (1-p)^x p \quad x = 0, 1, \ldots$$

and $A = \{0, 1, \ldots\}$.

Negative Binomial $X \sim NB(k, p)$ where X represents the number of successes in k $BERN(p)$ trials. We skip this for now.

Some examples of **continuous** r.v.s

Normal/Gaussian $X \sim N(\mu, \sigma^2)$ for $\mu \in \mathbb{R}$, $\sigma > 0$.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in \mathbb{R}$$

Gamma $X \sim GAM(\alpha, \beta)$ for $\alpha, \beta > 0$. The pdf may be left or right skewed.

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) \quad x \in \mathbb{R}^+$$

Note that the Gamma function Γ is defined as

$$\begin{aligned}
\Gamma(\alpha) &= (\alpha-1)\Gamma(\alpha-1) \quad \alpha > 1 \\
\Gamma(n) &= (n-1)! \\
\Gamma\left(\frac{1}{2}\right) &= \sqrt{\pi}
\end{aligned}$$

Exponential $X \sim EXP(\theta)$ for $\theta > 0$.

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \quad x \geq 0$$

Note that $EXP(\theta)$ is simply $GAM(1, \theta)$.

2 September 10, 2018

2.1 Cumulative distribution function (cdf)

We denote the *cumulative distribution function* (cdf) as $F(x) = P[X \leq x]$ with properties:

1. non-decreasing i.e. $F(a) \leq F(b)$ if $a \leq b$
2. $$\lim_{x \to -\infty} F(x) = 0$$
3. $$\lim_{x \to \infty} F(x) = 1$$
4. right-continuous, i.e. $\lim_{x \downarrow x_0} = F(x_0)$ (where $x \downarrow x_0$ denotes x approaches x_0 from x_0's right-hand side or in this case from above).

   Remark 2.1. If X is a continuous r.v. then $F(x)$ is also left-continuous i.e. $F(x)$ is continuous.

2.2 Location parameters

Example 2.1. If $X \sim N(\mu, 1)$, $\mu \in \mathbb{R}$, then μ is a location parameter for X where

$$f(x; \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} \quad x \in \mathbb{R}$$

$f(x, \mu)$ is *NOT completely specified* as $f(\cdot, \mu)$ cannot be calculated at x as μ is *unknown* (we would need to perform *statistical inference* to estimate μ).
On the other hand, $f(x; 0)$ is completely specified. Notice that

$$\begin{aligned} f(x; \mu) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu-0)^2}{2}} \\ &= f(x - \mu; 0) \end{aligned}$$

That is: the uncompletely specified $f(x; \mu)$ can be rewritten as a completely specified $f(\cdot; 0)$ evaluated at $x - \mu$.
μ is a *location parameter* for $X \sim N(\mu, 1)$.

Definition 2.1. A quantity η is a **location parameter** for X with a pdf $f(x; \eta)$ if

$$f(x; \eta) = f(x - \eta; 0)$$

Incerasing the value of the location parameter of the pdf shifts it to the right (e.g. for $N(\mu, 1)$).

For a continuous r.v. X with a location parameter η

$$\begin{aligned} F(x; \eta) &= P[X \leq x; \eta] \\ &= \int_{-\infty}^{x} f(t; \eta)\, \mathrm{d}t \\ &= \int_{-\infty}^{x} f(t - \eta; 0)\, \mathrm{d}t \end{aligned}$$

since η is a location parameter for our pdf f. Let $s = t - \eta$, then

$$\begin{aligned} &= \int_{-\infty}^{x-\eta} f(s;0)\,\mathrm{d}s \\ &= F(x-\eta;0) \end{aligned}$$

Therefore η is a location parameter *iff* $F(x;\eta) = F(x-\eta;0)$.

2.3 Scale parameters

Example 2.2. Let $X \sim EXP(\theta)$, $\theta > 0$ (as we will see, θ is a scale parameter for X). Recall

$$f(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}} \quad \theta > 0$$

is *NOT completely specified* as θ is unknown.
However $f(x;1) = \exp(-x)$ for $x > 0$ is the pdf of $EXP(1)$ which is completely satisfied. Note that

$$f(x;\theta) = \frac{1}{\theta}\exp(-\frac{x}{\theta}) = \frac{1}{\theta}f(\frac{x}{\theta};1)$$

θ is a *scale parameter* for $X \sim EXP(\theta)$, $\theta > 0$.

Definition 2.2. A quantity θ is a **scale parameter** if its pdf satisfies

$$f(x;\theta) = \frac{1}{\theta}f(\frac{x}{\theta};1) \quad \theta > 0$$

That is: the uncompletely specified pdf can be re-written as the product of $\frac{1}{\theta}$ and a completely specified pdf $f(\cdot;1)$ evaluated at $\frac{x}{\theta}$.

How about the corresponding cdf (for a continuous r.v awith scale parameter θ)?

$$\begin{aligned} F(x;\theta) &= \int_{-\infty}^{x} f(t;\theta)\,\mathrm{d}t \\ &= \int_{-\infty}^{x} f(\frac{t}{\theta};1)\frac{1}{\theta}\,\mathrm{d}t \end{aligned}$$

since θ is a scale parameter. Let $s = \frac{t}{\theta}$ (so $\mathrm{d}s = \frac{\mathrm{d}t}{\theta}$), thus

$$\begin{aligned} &= \int_{-\infty}^{\frac{x}{\theta}} f(s;1)\,\mathrm{d}s \\ &= F(\frac{x}{\theta};1) \end{aligned}$$

Therefore θ is a scale parameter *iff* $F(x;\theta) = F(\frac{x}{\theta};1)$.

2.4 Pivotal quantities

Remark 2.2. If η is a location parameter, then $\hat{\eta} - \eta$ is a pivotal quantity for constructing a confidence interval for η (where $\hat{\eta}$ is the Maximum Likelihood Estimate (MLE) of η).
If θ is a scale parameter, then $\frac{\hat{\theta}}{\theta}$ is a pivotal quantity for construct a confidence interval for θ.

3 September 12, 2018

3.1 Pdf of a function

We want to find the pdf of a function of one r.v.

Method 1 Let $Y = h(x)$. If $h(\cdot)$ is **a 1-1 function** then $h(\cdot)$ is either strictly increasing or strictly decreasing.

1. When $h(\cdot)$ is strictly increasing ($h^{-1}(\cdot)$ exists and is also strictly increasing): let $G(y)$ be the cdf of Y and $g(y)$ be the pdf of Y.
   Given that X is a continuous r.v. with pdf $f(x)$ and cdf $F(x)$, then

$$G(y) = P[Y \leq y] = P[h(X) \leq y] = P[X \leq h^{-1}(y)] = F(h^{-1}(y))$$

   For the pdf $g(y)$, we have

$$\begin{aligned} g(y) = \frac{dG(y)}{dy} &= \frac{dF(h^{-1}(y))}{dy} \\ &= f(h^{-1}(y)) \cdot \frac{\partial h^{-1}(y)}{\partial y} \\ &= f(h^{-1}(y)) \cdot \left| \frac{\partial h^{-1}(y)}{\partial y} \right| \end{aligned}$$

   since $h^{-1}(\cdot)$ is strictly increasing, we have $\frac{\partial h^{-1}(y)}{\partial y} > 0$ (so we can add an absolute sign).

2. When $h(\cdot)$ and thus $h^{-1}(\cdot)$ is strictly decreasing we have

$$\begin{aligned} G(y) = P[h(X) \leq y] &= P[h^{-1}(h(X)) \geq h^{-1}(y)] \\ &= P[X \geq h^{-1}(y)] \\ &= 1 - P[X < h^{-1}(y)] \\ &= 1 - P[X \leq h^{-1}(y)] \qquad P[X = h^{-1}(y)] = 0 \text{ since X is continuous} \\ &= 1 - F(h^{-1}(y)) \end{aligned}$$

   For the pdf $g(y)$

$$\begin{aligned} g(y) = \frac{dG(y)}{dy} &= \frac{d1 - F(h^{-1}(y))}{dy} \\ &= -f(h^{-1}(y)) \cdot \frac{\partial h^{-1}(y)}{\partial y} \\ &= f(h^{-1}(y)) \cdot \left| \frac{\partial h^{-1}(y)}{\partial y} \right| \end{aligned}$$

   since $h^{-1}(\cdot)$ is strictly decreasing thus $\frac{\partial h^{-1}(y)}{\partial y} < 0$, hence the absolute sign.

So if $h(\cdot)$ is a **1-1 function**, we have for $Y = h(X)$ the pdf

$$g(y) = f(h^{-1}(y)) \cdot \left| \frac{\partial h^{-1}(y)}{\partial y} \right|$$

How do we find the support set for Y? Let A be the support set of X and B be the support set for Y. Let $h : A \to B^*$ where B^* is the image of A under $h(\cdot)$.

Thus we have $B = \{y \mid y \in B^* \text{ and } g(y) > 0\}$.

Example 3.1. Let X have a pdf $f(x) = \frac{\theta}{x^{\theta+1}}$ where $x \geq 1$ and $\theta > 0$.

Find the pdf of $Y = \log X$ (natural log).

We have $h(X) = \log X$ thus $X = e^Y = h^{-1}(Y)$. Since $h(x)$ is 1-1 we can use our previous result:

$$f(h^{-1}(y) = f(e^y) = \frac{\theta}{(e^y)^{\theta+1}}$$

Also

$$\frac{\partial h^{-1}(y)}{\partial y} = \frac{\partial e^y}{\partial y} = e^y$$

Thus we have

$$\begin{aligned} g(y) &= \frac{\theta}{e^{y\theta} e^y} \cdot |e^y| \\ &= \frac{\theta}{e^{y\theta} e^y} \cdot e^y \\ &= \frac{\theta}{e^{y\theta}} \end{aligned}$$

To find the support, note that $h(x) = \log X$ has support $A = \{x \mid x \geq 1\}$ thus $h : A \to B^* = \{y \mid y \geq 0\}$. Note that $g(y) = \frac{\theta}{e^{y\theta}} > 0$ for all $y \in \mathbb{R}$, thus the support for Y is $B = B^* = \{y \mid y \geq 0\}$.

Method 2 For functions $h(\cdot)$ that are not 1-1, we use the cdf technique.

Example 3.2. Let $X \sim N(0,1)$ and $Y = X^2$: find the pdf $G(Y)$ of Y.

$$G(y) = P[Y \leq y] = P[X^2 \leq y]$$

Note that $P[X^2 \leq 0] = P[X^2 = 0] = 0$ since $x^2 \geq 0$ for all $x \in \mathbb{R}$, so if $y = 0$ then $G(y) = 0$.

For $y > 0$, we have

$$\begin{aligned} G(y) &= P[X^2 \leq y] \\ &= P[-\sqrt{y} \leq X \leq \sqrt{y}] \\ &= 2P[0 \leq X \leq \sqrt{y}] && N(0,1) \text{ is symmetric} \\ &= 2\int_0^{\sqrt{y}} f(x)\,\mathrm{d}x \\ &= 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathrm{d}x \end{aligned}$$

We require $g(y) = \frac{dG(y)}{dy}$.

From Fundamental Theorem of Calculus, if $f(x)$ is cont. on $[a,b]$ and $g(x) = \int_a^x f(t)\,\mathrm{d}t\ \forall x \in [a,b]$ is cont. on $[a,b]$ then

$$\frac{dg(x)}{dx} = f(x) \quad \forall x \in [a,b]$$

Thus for all $y > 0$ we have

$$\begin{aligned}\frac{dG(y)}{dy} &= \frac{2}{\sqrt{2\pi}} \frac{d\int_0^{\sqrt{y}} e^{-\frac{x^2}{2}}\,\mathrm{d}x}{dy} \\ &= \frac{2}{\sqrt{2\pi}} e^{-\frac{(\sqrt{y})^2}{2}} \cdot \frac{d\sqrt{y}}{dy} \\ &= -\frac{1}{\sqrt{\pi y}} e^{-\frac{y}{2}}\end{aligned}$$

So $g(y) = \frac{1}{\sqrt{\pi y}} e^{-\frac{y}{2}}$ is the pdf of $Y \sim X^2(1)$

Note that $h : A \to B^*$ where $A = \mathbb{R}$, thus $B^* = \{y \mid y > 0\}$.

The support set of Y is B where $B = \{y \mid y \in B^* \text{ and } g(y) > 0\}$.

Notice that $G(y) = 0$ if $y = 0$ and $G(y)$ is not differentiable at $y = 0$, thus $g(0) = 0$ so $B = \{y \mid y > 0\}$.

4 September 14, 2018

4.1 Expectations

The expection $E(X)$ of a r.v. X exists if $E(|X|) < \infty$. It is defined as

Discrete r.v. X

$$E(X) = \sum_{x \in A} x \cdot f(x)$$

By the Law of the Unconscious Statistician (LOTUS)

$$E(h(X)) = \sum_{x \in A} h(x) \cdot f(x)$$

Continuous r.v. X

$$\begin{aligned}E(X) &= \int_A x f(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x\end{aligned}$$

LOTUS holds for continuous r.v.'s as well

$$E(h(X)) = \int_A h(x) \cdot f(x)\,\mathrm{d}x$$

4.2 Markov's inequality

Theorem 4.1 (Markov's inequality)**.** Markov's inequality states that

$$P[|X| \geq c] \leq \frac{E[|X|^k]}{c^k}$$

for all $c, k > 0$.

Proof. Note that $P[|X| \geq c] = P[X \leq -c] + P[X \geq c]$ or the tail probabilities beyond $-c$ and c.
Thus Markov's inequality gives an *upper bound* for the tail probabilities.
In the countinuous case we have for the RHS

$$P[|X| \geq c] = \int_{\{x||x| \geq c\}} f(x)\,\mathrm{d}x$$

For the LHS we have

$$\begin{aligned}
\frac{E[|X|^k]}{c^k} = E[|\frac{X}{c}|^k] &= \int_{-\infty}^{\infty} |\frac{x}{c}|^k f(x)\,\mathrm{d}x \\
&= \int_{x||x|\geq c} |\frac{x}{c}|^k f(x)\,\mathrm{d}x + \int_{x||x|<c} |\frac{x}{c}|^k f(x)\,\mathrm{d}x \\
&\geq \int_{x||x|\geq c} |\frac{x}{c}|^k f(x)\,\mathrm{d}x && \text{right term is integral over non-negative function} \\
&\geq \int_{x||x|\geq c} f(x)\,\mathrm{d}x && |x| \geq c \Rightarrow |\frac{x}{c}|^k \geq 1
\end{aligned}$$

and the result follows. □

Example 4.1. Given $X \sim N(0, \sigma^2)$, what is a bound on $P[|X| \geq 3\sigma]$?
From Markov's inequality, let $k = 2$ (where $E[X^2] = \sigma^2$)

$$\begin{aligned}
P[|X| \geq 3\sigma] &\leq \frac{E[|X|^k]}{(3\sigma)^k} \\
&= \frac{E[X^2]}{9\sigma^2} \\
&= \frac{\sigma^2}{9\sigma^2} \\
&= \frac{1}{9}
\end{aligned}$$

Since $P[|X| \geq 3\sigma] \leq \frac{1}{9}$ then $P[|X| \leq 3\sigma] \geq 1 - \frac{1}{9} = \frac{8}{9}$.
Thus X stays 3σ distance from 0 with a high probability of at least $\frac{8}{9}$.

4.3 Moment generating function (mgf)

Definition 4.1 (Moment generating function)**.** For a r.v. X the expectation

$$M_X(t) = E[e^{tX}]$$

is called the moment generating function (if the expectation exists).
One must state the values of t such that $M_X(t)$ exists ("domain of convergence").

Example 4.2. Let $X \sim GAM(\alpha, \beta)$, $\alpha, \beta > 0$. Find $M_X(t)$.

$$
\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= \int_0^\infty e^{tx} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \, \mathrm{d}x \\
&= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-x(\frac{1}{\beta}-t)} \, \mathrm{d}x
\end{aligned}
$$

Note that for any pdf $f(x)$ we have $\int_A f(x)\,\mathrm{d}x = 1$, thus $\int_0^\infty \frac{1}{\beta^\alpha\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} = 1$ thus $\int_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}\,\mathrm{d}x = \beta^\alpha\Gamma(\alpha)$. Thus we have from before

$$
\begin{aligned}
\frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-x(\frac{1}{\beta}-t)} \, \mathrm{d}x &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \left(\frac{1}{\frac{1}{\beta} - t} \right)^\alpha \Gamma(\alpha) \\
&= \frac{1}{(1-\beta t)^\alpha}
\end{aligned}
$$

when $(\frac{1}{\beta} - t)^{-1} > 0$ i.e. $t < \frac{1}{\beta}$. What if $t \geq \frac{1}{\beta}$? When $t = \frac{1}{\beta}$ our integral becomes $\int_{-\infty}^\infty x^{\alpha-1}\,\mathrm{d}x$ which goes to infinity for $\alpha > 0$.
Similarly it goes to infinity when $t > \frac{1}{\beta}$.

5 September 17, 2018 and September 19, 2018

5.1 Derivatives of mgf

For the continuous case (similarly for discrete) we can take the derivative of the mgf $M_X(t)$

$$
\begin{aligned}
\frac{dM_X(t)}{dt} &= \frac{d}{dt} \sum_{-\infty}^{\infty} e^{tX} f(x) \, \mathrm{d}x \\
&= \sum_{-\infty}^{\infty} \frac{d}{dt} [e^{tX} f(x)] \, \mathrm{d}x && \text{Leibniz rule} \\
&= \sum_{-\infty}^{\infty} x e^{tX} f(x) \, \mathrm{d}x
\end{aligned}
$$

We can claerly see when $t = 0$ we have the expected value $E[X]$. Similarly

$$
\begin{aligned}
\frac{\mathrm{d}^2 M_X(t)}{\mathrm{d}t^2} &= \frac{d}{dt} \big[\frac{d}{dt} M_X(t) \big] \\
&= \frac{d}{dt} \big[\sum_{-\infty}^{\infty} x e^{tX} f(x) \, \mathrm{d}x \big] \\
&= \sum_{-\infty}^{\infty} \frac{d}{dt} [x e^{tX} f(x)] \, \mathrm{d}x \\
&= \sum_{-\infty}^{\infty} x^2 e^{tX} f(x) \, \mathrm{d}x
\end{aligned}
$$

which we recognize when $t = 0$ as the second moment $E[X^2]$.
In summary

$$\frac{\mathrm{d}^r}{\mathrm{d}t^r}M_X(t) = \int_{-\infty}^{\infty} x^r e^{tX} f(x)\,\mathrm{d}x \qquad r = 1, 2, \ldots$$

where

$$\begin{aligned}\Big(\frac{\mathrm{d}^r}{\mathrm{d}t^r}M_X(t)\Big)\Big|_{t=0} &= \Big(\int_{-\infty}^{\infty} x^r e^{tX} f(x)\,\mathrm{d}x\Big)\Big|_{t=0} \\ &= \int_{-\infty}^{\infty} x^r f(x)\,\mathrm{d}x \\ &= E[X^r]\end{aligned}$$

Example 5.1. For $X \sim GAM(\alpha, \beta)$ we have $M_X(t) = \frac{1}{(1-\beta t)^\alpha}$, $t < \frac{1}{\beta}$. Find $E[X]$ and $Var(X)$.
Note that $Var(X) = E[X^2] - E[X]^2$. Also

$$\begin{aligned}\frac{d}{dt}M_X(t) &= \frac{d}{dt}[(1-\beta t)^{-\alpha}] \\ &= (-\alpha)(-\beta)(1-\beta t)^{-\alpha-1} \\ &= \alpha\beta(1-\beta t)^{-\alpha-1}\end{aligned}$$

Thus $E[X] = \alpha\beta(1-\beta 0)^{-\alpha-1} = \alpha\beta$.
Similarly $E[X^2] = \alpha(\alpha+1)\beta^2$ thus $Var(X) = \alpha\beta^2$.

5.2 Joint cdf and pdf

The joint cdf $F[x, y]$ is defined as $P[X \le x \text{ and } Y \le y]$ or simply $P[X \le x, Y \le y]$.
Recall that for the cdf $F(x)$ for X, we have

1. $F(a) \le F(b)$ if $a \le b$
2. $\lim_{x\to-\infty} F(x) = 0$
3. $\lim_{x\to\infty} F(x) = 1$
4. $\lim_{x\downarrow x_0} F(x) = F(x_0)$ (right continuous)

Similarly, the properties for the *joint cdf* of X and Y are

1. For every fixed y, $F(x, y)$ is non-decreasing for x. Similarly for fixed x, $F(x, y)$ is non-decreasing for y.
2. For every fixed y, $\lim_{x\to-\infty} F(x, y) = 0$ (similarly with fixed x and $y \to -\infty$).
3. $\lim_{x,y\to\infty} F(x, y) = 1$
4. $$\begin{aligned}F_1(x) &= P[X \le x] = \lim_{y\to\infty} F(x, y) \\ F_2(y) &= P[Y \le y] = \lim_{x\to\infty} F(x, y)\end{aligned}$$

Comparing discrete and continuous joint r.v.s

Discrete r.v. For the pmf we have

$$f(x, y) = P(X = x, Y = y)$$

Our support set is $A = \{(x, y) \mid f(x, y) > 0\}$.

For the pmf, we have

1. $f(x, y) \geq 0$ for all $(x, y) \in \mathbb{R} \times \mathbb{R}$
2. $\sum \sum f(x, y) = 1$ where $(x, y) \in A$

To compute the marginal pmf for x we take

$$f_1(x) = \sum_{y \in \mathbb{R}} f(x, y)$$

(similarly for the marginal pmf for y).

Continuous r.v. For the pdf we have

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

Our support set is $A = \{(x, y) \mid f(x, y) > 0\}$.

For the pdf, we have

1. $f(x, y) \geq 0$ for all $(x, y) \in \mathbb{R} \times \mathbb{R}$
2. $\int \int f(x, y) \,\mathrm{d}x \,\mathrm{d}y = 1$ where $(x, y) \in A$

To compute the marginal pdf for x we take

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) \,\mathrm{d}y$$

(similarly for the marginal pmf for y).

Example 5.2. Suppose that X and Y are cont. r.v.s with joint pdf $f(x, y) = x + y$ for $0 < x < 1$ and $0 < y < 1$. Find

1. $P[X \leq \frac{1}{3}, Y \leq \frac{1}{2}] = F(\frac{1}{3}, \frac{1}{2})$
2. $P[X \leq Y]$
3. $P[X + Y \leq \frac{1}{2}]$
4. $P[XY \leq \frac{1}{2}]$
5. $f_1(x)$
6. $F(x, y)$
7. $F_1(x)$

Solution. Note that while we may be finding $P[X \leq \frac{1}{3}]$ which is generally everything to the right of $x = \frac{1}{3}$, we only want the region intersected by our support set. This is represented as the shaded region in the diagrams below.

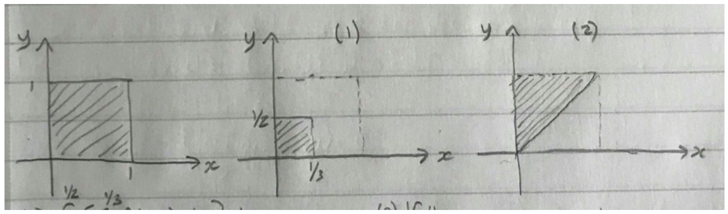


Figure 5.1: Diagram of area we are trying to integrate over for (1) and (2).

1. We sum over the shaded square area

$$\begin{aligned}
\int_0^{\frac{1}{2}} \left(\int_0^{\frac{1}{3}} f(x,y)\,\mathrm{d}x \right) \mathrm{d}y &= \int_0^{\frac{1}{2}} \left(\int_0^{\frac{1}{3}} x + y\,\mathrm{d}x \right) \mathrm{d}y \\
&= \int_0^{\frac{1}{2}} \left(\frac{x^2}{2} + xy \Big|_{x=0}^{x=1/3} \right) \mathrm{d}y \\
&= \int_0^{\frac{1}{2}} \frac{1}{18} + \frac{y}{3}\,\mathrm{d}y \\
&= \frac{y}{18} + \frac{y^2}{6} \Big|_{y=0}^{y=1/2} \\
&= \frac{5}{72}
\end{aligned}$$

2. If the region is not rectangular we pick one variable first, say y, and range from its smallest value to the largest value in its region.

   We then find the range of the other variable (x in this case) for every given y.

$$\begin{aligned}
P[X \leq Y] &= \int_0^1 \left(\int_0^y f(x,y)\,\mathrm{d}x \right) \mathrm{d}y \text{ OR} \\
&= \int_0^1 \left(\int_x^1 f(x,y)\,\mathrm{d}y \right) \mathrm{d}x
\end{aligned}$$

   We have

$$\begin{aligned}
P[X \leq Y] &= \int_0^1 \left(\int_0^y f(x,y)\,\mathrm{d}x \right) \mathrm{d}y \\
&= \int_0^1 \left(\int_0^y x + y\,\mathrm{d}x \right) \mathrm{d}y \\
&= \int_0^1 \frac{3y^2}{2}\,\mathrm{d}y \\
&= \frac{1}{2}
\end{aligned}$$

3. The region is the triangle under the line $y = \frac{1}{2} - x$ in quadrant 1.

$$\begin{aligned}
P[X+Y \leq \frac{1}{2}] &= P[Y \leq -x + \frac{1}{2}] \\
&= \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}-y} x + y \, \mathrm{d}x \, \mathrm{d}y \\
&= \int_0^{\frac{1}{2}} \frac{x^2}{2} + xy \Big|_{x=0}^{x=\frac{1}{2}-y} \mathrm{d}y \\
&\vdots \\
&= \frac{1}{24}
\end{aligned}$$

4. We have 1 - $XY \geq \frac{1}{2}$ thus

$$\begin{aligned}
P[XY \geq \frac{1}{2}] &= \int_{\frac{1}{2}}^{1} \int_{\frac{1}{2y}}^{1} f(x,y) \, \mathrm{d}x \, \mathrm{d}y \\
&= \frac{1}{4}
\end{aligned}$$

Thus $P[XY \leq \frac{1}{2}] = \frac{3}{4}$.
Otherwise we would need to break it apart in two parts (when $y \leq \frac{1}{2}$ and when $y > \frac{1}{2}$):

$$\begin{aligned}
P[XY \leq \frac{1}{2}] &= \int_{\frac{1}{2}}^{1} \int_0^{\frac{1}{2y}} f(x,y) \, \mathrm{d}x \, \mathrm{d}y + \int_0^{\frac{1}{2}} \int_0^1 f(x,y) \, \mathrm{d}x \, \mathrm{d}y \\
&= \frac{3}{4}
\end{aligned}$$

5. We have

$$\begin{aligned}
f_1(x) &= \int_{-\infty} \infty f(x,y) \, \mathrm{d}y \\
&= \int_0^1 f(x,y) \, \mathrm{d}y && A = \{(x,y) \mid 0 < x < 1, 0 < y < 1\} \\
&= \int_0^1 x + y \, \mathrm{d}y \\
&= x + \frac{1}{2} \qquad 0 < x < 1
\end{aligned}$$

Similarly $f_2(y) = \int_0^1 f(x,y) \, \mathrm{d}x = y + \frac{1}{2}$ for $0 < y < 1$.

6. If $x \leq 0$ **or** $y \leq 0$, then $F(x,y) = 0$.
Similarly if $x \geq 1$ and $y \geq 1$, then $F(x,y) = 1$.

If $0 < x \leq 1$ and $0 < y \leq 1$

$$\begin{aligned}F(x,y) &= \int_0^y \int_0^x f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{2}x^2y + \frac{1}{2}xy^2\end{aligned}$$

If $0 < x \leq 1$ and $y > 1$

$$\begin{aligned}F(x,y) &= P[X \leq x, Y \leq y] \\ &= P[X \leq x, Y \leq 1] \\ &= F(x,1) \\ &= \frac{1}{2}(x^2 + x)\end{aligned}$$

Similarly for $x > 1$ and $0 < y \leq 1$, $F(x,y) = \frac{1}{2}(y^2 + y)$.

7. Note that $F_1(x) = \lim_{y\to\infty} F(x,y)$. From above we have

$$F_1(x) = \begin{cases} \lim_{y\to\infty} F(x,y) = \lim_{y\to\infty} 0 = 0 & x \leq 0 \\ \lim_{y\to\infty} F(x,y) = \lim_{y\to\infty} 1 = 1 & x \geq 1 \\ \lim_{y\to\infty} F(x,y) = \lim_{y\to\infty} \frac{1}{2}(x^2 + x) & x0 < x < 1 \end{cases}$$

6 September 21, 2018

6.1 Independence

X and Y are independent if $P[X \in A, Y \in B] = P[X \in A]P[Y \in B]$ for any $A, B \subseteq \mathbb{R}$.

Corollary 6.1. If X and Y are independent, then $h(X)$ and $g(Y)$ are independent for any real-valued functions $h(\cdot)$ and $g(\cdot)$.

Proof. To show $h(X), g(Y)$ are independent, we need to prove

$$P[h(X) \in A^*, g(Y) \in B^*] = P[h(X) \in A^*]P(g(Y) \in B^*]$$

for any $A^*, B^* \subseteq \mathbb{R}$.
Note that for functions $h : A \to A^*$ and $g : B \to B^*$, $x \in A \iff h(x) \in A^*$ and similarly $y \in B \iff g(x) \in B^*$.
Thus $P[h(X) \in A^*, g(Y) \in B^*] = P[X \in A, Y \in B] = P[X \in A]P[Y \in B]$ as X, Y are independent.
Again since we have an $\iff$ correspondence we have $P[h(X) \in A^*]P[g(Y) \in B^*]$. □

Theorem 6.1. X, Y are independent **if and only if** *either*

$$f(x,y) = f_1(x)f_2(y) \qquad \forall (x,y) \in A_1 \times A_2$$

where A_1, A_2 are the support sets for X and Y, respectively, *OR*

$$F(x,y) = F_1(x)F_2(y)\forall (x,y) \in \mathbb{R} \times \mathbb{R}$$

Example 6.1. For $f(x,y) = x + y$, $0 < x < 1$, $0 < y < 1$, are X, Y independent? Why?
Note that from before we found that $f_1(x) = \frac{1}{2} + x$ for $0 < x < 1$; $f_2(y) = \frac{1}{2} + y$ for $0 < y < 1$.

Does $f(x,y) = f_1(x)f_2(y)$ for all $(x,y) \in A_1 \times A_2$, where $A_1 = \{x \mid 0 < x < 1\}, A_2 = \{y \mid 0 < y < 1\}$.
No: since $(x+y) \neq (\frac{1}{2}+x)(\frac{1}{2}+y)$ for all $(x,y) \in (0,1) \times (0,1)$, thus X, Y are not independent.

6.2 Factorization independence theorem

Theorem 6.2 (Factorization independence)**.** Suppose X, Y have joint pdf $f(x,y)$ and support set $A = \{(x,y) \mid f(x,y) > 0\}$.
Then X, Y are independent **if and only if** $A = A_1 \times A_2$ and $f(x,y) = h(x) \cdot g(y)$ for some non-negative functions $h(\cdot)$ and $g(\cdot)$ for all $(x,y) \in A$.

Remark 6.1. We need to check that

1. $A = A_1 \times A_2$ i.e. A is rectangular (otherwise we would have undefined values for $f(x,y)$ for some $x \in A_1$ or $y \in A_2$).
2. Check $f(x,y) = h(x) \cdot g(y)$

Example 6.2. Suppose X, Y have joint pdf

$$f(x,y) = \frac{\theta^{x+y}e^{-2\theta}}{x!y!} \qquad x,y = 0,1,2,\ldots$$

Are X, Y independent? Why?

1. Does $A = A_1 \times A_2$? Yes since we have $A_1 = \{x \mid x = 0,1,2,\ldots\}$ and $A_2 = \{y \mid y = 0,1,2,\ldots\}$.
2. We see that
$$f(x,y) = (\frac{\theta^x}{x!})(\frac{\theta^y e^{-2\theta}}{y!})$$
and there are many other functions where each function has complementary constant scaling factors.

   Remark 6.2. Note that $h(\cdot)$ and $g(\cdot)$ may not be true pdfs (i.e. they may not sum up to 1 over the support set: see the remark below).

Thus X, Y are independent by the Factorization theorem.

Remark 6.3. When the Factorization theorem holds, $h(x)$ is *proportional* to $f_1(x)$ and $g(y)$ is proportional to $f_2(y)$.

Proof. We have

$$\begin{aligned} f_1(x) &= \sum_{y=0}^{\infty} f(x,y) \\ &= \sum_{y=0}^{\infty} h(x)g(y) \\ &= h(x)\sum_{y=0}^{\infty} g(y) \end{aligned}$$

From the example above, we had $g(y) = \frac{\theta^y}{y!}$, so

$$f_1(x) = h(x) \sum_{y=0}^{\infty} \frac{\theta^y}{y!} = e^{\theta} h(x) = \frac{\theta^x e^{-\theta}}{x!}$$

Thus $X \sim POI(\theta)$ and similarly $Y \sim POI(\theta)$. $\square$

Example 6.3. Suppose X, Y have joint pdf

$$f(x, y) = \frac{2}{\pi} \qquad 0 \leq x \leq \sqrt{1 - y^2}, -1 \leq y \leq 1$$

Are X, Y independent? Why?
Note that $A \neq A_1 \times A_2$ since we have $A_1 = \{x \mid 0 \leq x \leq 1\}$ and $A_2 = \{y \mid -1 \leq y \leq 1\}$.
Since A does not have the support set that is the recetangular bounds of $A_1 \times A_2$ there is no way to factorize our joint pdf into the product of two marginal pdfs.

7 September 24, 2018

7.1 Conditional pmf/pdf

Definition 7.1. We define the **conditional pmf/pdf** of x on y to be

$$f_1(x \mid y) = \frac{f(x, y)}{f_2(y)} \qquad (x, y) \in A \text{ and } f_2(y) \neq 0$$

where A is the support set for (X, Y) (i.e. $f(x, y)$)

Properties of $f_1(x \mid y)$ for discrete and continuous r.v's:

Discrete r.v.s
1. $f_1(x \mid y) \geq 0$ for all $(x, y) \in \mathbb{R} \times \mathbb{R}$
2. $\sum_{x \in \mathbb{R}} f_1(x \mid y) = 1$

Continuous r.v.s
1. $f_1(x \mid y) \geq 0$ for all $(x, y) \in \mathbb{R} \times \mathbb{R}$
2. $\int_{-\infty}^{\infty} f_1(x \mid y) \, \mathrm{d}x = 1$

Similarly $f_2(y \mid x) = \frac{f(x,y)}{f_1(x)}$ and $f_1(x) \neq 0$.

7.2 Product rule

The product rules states that

$$\begin{aligned} f(x, y) &= f_1(x \mid y) f_2(y) \qquad \text{OR} \\ &= f_2(y \mid x) f_1(x) \end{aligned}$$

Application of product rule: if $f_1(x \mid y)$ and $f_2(y)$ are given, we can find $f_1(x)$ (Take $\int_{y \in A} f_1(x \mid y) f_2(y) \, \mathrm{d}y$ in the continuous case).

Example 7.1. Let $Y \sim POI(\mu)$ and $X \mid Y = y \sim BIN(y, p)$. Find the marginal distribution of X.
We will take the route $f_1(x \mid y)$ and $f_2(y) \to f(x, y) \to f_1(x)$.

Note that

$$f_2(y) = \frac{\mu^y e^{-\mu}}{y!} \qquad y = 0, 1, 2, \ldots$$

Also

$$f_1(x \mid y) = \binom{y}{x} p^x (1-p)^{y-x} \qquad x = 0, 1, \ldots, y$$

Thus we have

$$\begin{aligned} f(x,y) = f_1(x \mid y) f_2(y) &= \frac{\mu^y e^{-\mu}}{y!} \cdot \frac{y!}{(y-x)!x!} p^x (1-p)^{y-x} \\ &= \frac{\mu^y e^{-\mu}}{(y-x)!x!} p^x (1-p)^{y-x} \end{aligned}$$

where $x = 0, 1, \ldots, y$ and $y = 0, 1, \ldots$ i.e. $0 \le x \le y$ (and $y \ge 0$). We need to be aware of these bounds when marginalizing over x, so

$$\begin{aligned} f_1(x) &= \sum_{y=x}^{\infty} f(x,y) \\ &= \frac{e^{-\mu} p^x (1-p)^{-x}}{x!} \sum_{y=x}^{\infty} \frac{\mu^y (1-p)^y}{(y-x)!} \\ &= \frac{e^{-\mu} \left(\frac{p}{1-p}\right)^x}{x!} \sum_{y=x}^{\infty} \frac{(\mu(1-p))^y}{(y-x)!} \\ &= \frac{e^{-\mu} \left(\frac{p}{1-p}\right)^x (\mu(1-p))^x}{x!} \sum_{y=x}^{\infty} \frac{(\mu(1-p))^{y-x}}{(y-x)!} \\ &= \frac{e^{-\mu} (\mu p)^x}{x!} \sum_{n=0}^{\infty} \frac{(\mu(1-p))^n}{n!} && n = y - x \\ &= \frac{e^{-\mu} (\mu p)^x}{x!} e^{\mu(1-p)} && \text{Taylor series of } e^{\mu(1-p)} \\ &= \frac{e^{-\mu p} (\mu p)^x}{x!} \qquad x = 0, 1, \ldots \end{aligned}$$

that is $X \sim POI(\mu p)$.

Example 7.2. let $Y \sim GAM(\alpha, 1)$ (not $GAM(\alpha, \frac{1}{\theta})$ in the notes) and $X \mid Y = y \sim WEI(y^{-\frac{1}{p}}, p)$ (Weibull distribution). Find the marginal pdf of X.
We will be following $f_1(x \mid y)$ and $f_2(y) \to f(x,y) \to f_1(x)$.
Note that

$$f_1(y) = \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y}$$

For a given $X \sim WEI(\theta, \beta)$ we have

$$f(x) = \frac{\beta}{\theta^\beta} x^{\beta-1} e^{-(\frac{x}{\theta})^\beta}$$

where $x > 0$. Thus we have

$$f_1(x \mid y) = \frac{p}{(y^{-(\frac{1}{p})^p}} x^{p-1} e^{-(\frac{x}{y^{-\frac{1}{p}}})^p}$$

Note we have $A = \{(x, y) \mid x > 0, y > 0\}$ thus

$$\begin{aligned} f_1(x) &= \int_0^\infty f(x,y)\,\mathrm{d}y \\ &= \frac{px^{p-1}}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y} y e^{-x^p y}\,\mathrm{d}y \\ &= \frac{px^{p-1}}{\Gamma(\alpha)} \int_0^\infty y^{\alpha} e^{-y(1+x^p)}\,\mathrm{d}y \end{aligned}$$

Recall we have

$$\begin{aligned} \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}\,\mathrm{d}x = 1 \\ \Rightarrow \Gamma(\alpha)\beta^\alpha = \int_0^\infty x^{\alpha-1} e^{-\frac{x}{\beta}}\,\mathrm{d}x \end{aligned}$$

So we have

$$\int_0^\infty y^\alpha e^{-y(1+x^p)}\,\mathrm{d}y = \Gamma(\alpha+1)[(1+x^p)^{-1}]^{\alpha+1}$$

Thus

$$\begin{aligned} f_1(x) &= \frac{px^{p-1}}{\Gamma(\alpha)} \Gamma(\alpha+1)[(1+x^p)^{-1}]^{\alpha+1} \\ &= p\alpha \cdot \frac{x^{p-1}}{(1+x^p)^{\alpha+1}} \end{aligned}$$

Note that $X \sim Burr(p, \alpha)$ or the Burr distribution.

7.3 Independence and condition pmf/pdfs

Note that $f(x, y) = f_1(x)f_2(y) = f_1(x \mid y)f_2(y)$, therefore X and Y are independent **if and only if** $f_1(x \mid y) = f_1(x)$ (or similarly if $f_2(y \mid x) = f_2(y)$).

Example 7.3. Let $f(x, y) = \frac{2}{\pi}$ where $0 \le x \le \sqrt{1-y^2}$, $-1 \le y \le 1$.
Note that

$$\begin{aligned} f_1(x) &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f(x,y)\,\mathrm{d}y = \frac{4\sqrt{1-x^2}}{\pi} && 0 < x < 1 \\ f_2(y) &= \int_0^{\sqrt{1-y^2}} f(x,y)\,\mathrm{d}x = \frac{2\sqrt{1-y^2}}{\pi} && -1 < y < 1 \\ f_1(x \mid y) &= \frac{f(x,y)}{f_2(y)} = \frac{1}{\sqrt{1-y^2}} && 0 \le x \le \sqrt{1-y^2}, -1 < y < 1 \end{aligned}$$

Since $f_1(x, y) \neq f_1(x)$ then X and Y are not independent.

8 September 26, 2018

8.1 Joint expectation

We define the **joint expectation** for discrete and continuous r.v.s:

Discrete The joint expectation is

$$E[h(X,Y)] = \sum_{x\in\mathbb{R}}\sum_{y\in\mathbb{R}} h(x,y)\cdot f(x,y)$$

Continuous The joint expectation is

$$E[h(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)\cdot f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

Theorem 8.1. If X, Y are independent, then

$$E[g(X)h(Y)] = E[g(X)]\cdot E[h(Y)]$$

for any real-valued functions $g(\cdot)$ and $h(\cdot)$.

Proof. Note that $g(X)$ and $h(Y)$ are functions of X and Y, thus by the joint expectation

$$\begin{aligned}
E[g(X)h(Y)] &= \int_{-\infty}^{\infty}\big[\int_{-\infty}^{\infty} g(x)h(y)f(x,y)\,\mathrm{d}x\big]\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\big[\int_{-\infty}^{\infty} g(x)h(y)f_1(x)f_2(y)\,\mathrm{d}x\big]\,\mathrm{d}y && \text{independence} \\
&= \big[\int_{-\infty}^{\infty} g(x)f_1(x)\,\mathrm{d}x\big]\big[\int_{-\infty}^{\infty} h(y)f_2(y)\,\mathrm{d}y\big] \\
&= E[g(X)]\cdot E[h(Y)]
\end{aligned}$$

□

8.2 Conditional expectation

We define the **conditional expectation** for discrete and continuous r.v.s:

Discrete The conditional expectation is

$$E[Y \mid X=x] = \sum_{y\in\mathbb{R}} y f_2(y \mid x)$$

by LOTUS

$$E[h(Y) \mid X=x] = \sum_{y\in\mathbb{R}} h(y) f_2(y \mid x)$$

Continuous The conditional expectation is

$$E[Y \mid X=x] = \int_{-\infty}^{\infty} y f_2(y \mid x)\,\mathrm{d}y$$

by LOTUS

$$E[h(Y) \mid X = x] = \int_{-\infty}^{\infty} h(y) f_2(y \mid x)\,\mathrm{d}y$$

Remark 8.1. 1. $E[Y \mid X = x]$ is a function of x only since we've summed over our support for Y.

2. If X, Y are independent, then $E[Y \mid X = x] = E[Y]$ since

$$\begin{aligned} E[Y \mid X = x] &= \int_{-\infty}^{\infty} y f_2(y \mid x)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} y f_2(y)\,\mathrm{d}y && \text{independence} \\ &= E[Y] \end{aligned}$$

similarly $E[h(Y) \mid X = x] = E[h(Y)]$.

Example 8.1. Let $f(x, y) = \frac{2}{\pi}$ where $0 \le x \le \sqrt{1 - y^2}$, $-1 \le y \le 1$.
Note that $A = \{(x, y) \mid 0 \le x \le \sqrt{1 - y^2}, -1 \le y \le 1\}$ or $A = \{(x, y) \mid 0 \le x \le 1, -\sqrt{1 - x^2} \le y \le \sqrt{1 - x^2}\}$, where $A_1 = \{x \mid 0 \le x \le 1\}$ and $A_2 = \{y \mid -1 \le y \le 1\}$.
Thus the conditional pdfs are

$$f_2(y \mid x) = \frac{f(x, y)}{f_1(x)} = \frac{1}{2\sqrt{1 - x^2}}$$

for $(x, y) \in A$ and $f_1(x) \neq 0$ thus $0 \le x < 1$ and $-\sqrt{1 - x^2} \le y \le \sqrt{1 - x^2}$.
Note that $Y \mid X = x$ is actually a uniform distribution symmetric around $y = 0$ ($UNIF(-\sqrt{1 - x^2}, \sqrt{1 - x^2})$ for $0 \le x < 1$), thus we expect $E[Y \mid X = x] = 0$. We verify

$$\begin{aligned} E[Y \mid X = x] &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y f_2(y \mid x)\,\mathrm{d}y \\ &= \frac{1}{2\sqrt{1 - x^2}} \left(\frac{1}{2} y^2 \Big|_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\right) \\ &= \frac{1}{2\sqrt{1 - x^2}} \cdot 0 \\ &= 0 \end{aligned}$$

We can also find $E[Y^2 \mid X = x]$

$$\begin{aligned} E[Y^2 \mid X = x] &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y^2 f_2(y \mid x)\,\mathrm{d}y \\ &= \frac{1}{2\sqrt{1 - x^2}} \left(\frac{1}{3} y^3 \Big|_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\right) \\ &= \frac{(1 - x^2)}{3} \qquad 0 \le x < 1 \end{aligned}$$

For $Var(Y \mid X = x)$ we have

$$\begin{aligned}
Var(Y \mid X = x) &= E[Y^2 \mid X = x] - E[Y \mid X = x]^2 \\
&= \frac{(1-x^2)}{3} - 0^2 \\
&= \frac{(1-x^2)}{3} \qquad 0 \leq x < 1
\end{aligned}$$

Remark 8.2. $E[Y \mid X = x]$ and $E[h(Y) \mid X = x]$ are functions of x, thus $E[Y \mid X]$ is a function of X (function of a random variable is a random variable).

8.3 Expectation of a conditional expectation

Theorem 8.2. We claim $E[E[h(Y) \mid X]] = E[h(Y)]$.
Let $g(X) = E[h(Y) \mid X]$, thus we have a function of X which from LOTUS we know

$$\begin{aligned}
E[g(X)] &= \int_{-\infty}^{\infty} g(x) f_1(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} E[h(Y) \mid X = x] f_1(x)\,\mathrm{d}x && g(x) = E[h(Y) \mid X = x] \\
&= \int_{-\infty}^{\infty} \big[\int_{-\infty}^{\infty} h(y) f_2(y \mid x)\,\mathrm{d}y \big] f_1(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \big[\int_{-\infty}^{\infty} h(y) f_2(y \mid x) f_1(x)\,\mathrm{d}y \big]\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \big[\int_{-\infty}^{\infty} h(y) f(x, y)\,\mathrm{d}y \big]\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \big[\int_{-\infty}^{\infty} h(y) f(x, y)\,\mathrm{d}x \big]\,\mathrm{d}y && \text{Fubini's theorem} \\
&= \int_{-\infty}^{\infty} h(y) \big[\int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}x \big]\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} h(y) f_2(y)\,\mathrm{d}y \\
&= E[h(Y)]
\end{aligned}$$

8.4 Variance as sum of conditional expectations

Theorem 8.3. We claim $Var(Y) = E[Var(Y \mid X)] + Var(E[Y \mid X])$.
We have $Var(Y) = E[Y^2] - E[Y]^2$ on the LHS.
On the RHS we have

$$\begin{aligned}
E[Var(Y \mid X)] + Var(E[Y \mid X]) &= E\big[E[Y^2 \mid X] - E[Y \mid X]^2\big] + \big(E[E[Y \mid X]^2] - E[E[Y \mid X]]^2\big) \\
&= E[E[Y^2 \mid X]] - E[E[Y \mid X]^2] + E[E[Y \mid X]^2] - E[E[Y \mid X]]^2 \\
&= E[E[Y^2 \mid X]] - E[E[Y \mid X]]^2 \\
&= E[Y^2] - E[Y]^2
\end{aligned}$$

where the last equality follows from $E[E[h(Y) \mid X]] = E[h(Y)]$.

Example 8.2. Suppose $Y \mid P = p \sim BIN(n, p)$ and $P \sim UNIF(0, 1)$. Find $E[Y]$ and $Var(Y)$.
Note that

$$E[Y] = E[E[Y \mid P]] = E[nP] = n \cdot E[P] = \frac{n}{2}$$

Similarly

$$\begin{aligned} Var(Y) = E[Var(Y \mid P)] + Var(E[Y \mid P]) &= E[nP(1-P)] + Var(nP) \\ &= nE[P] - nE[P^2] + n^2Var(P) \\ &= n\frac{1}{2} - n(Var(P) + E[P]^2) + n^2Var(P) \\ &= \frac{n}{2} - n(\frac{1}{12} + \frac{1}{2^2}) + n^2\frac{1}{12} \\ &= \frac{5n}{6} + \frac{n^2}{12} \end{aligned}$$

9 September 28, 2018

9.1 Joint moment generating function (mgf)

Recall the moment generating function (mgf) of X is defined as $M_X(t) = E[e^{tX}]$. For a given MGF:

1. State the values of t such that $M_X(t)$ exists, i.e. $E[e^{tX}] < \infty$.
2. Uniqueness: if X and Y have the same mgf, then X and Y are identically distributed (i.e. X, Y have the same pmf/pdf, cdf, etc.).

Definition 9.1 (Joint mgf)**.** The **joint mgf** of X and Y is defined as

$$M(t_1, t_2) = E[e^{(t_1,t_2)\cdot(X,Y)^T}] = E[e^{t_1X+t_2Y}] = E[e^{t_1X}e^{t_2Y}]$$

where one needs to state the values of t_1, t_2 such that $M(t_1, t_2)$ exists.

Note that given the joint mgf, it is straightforward to derive the marginal mgf

$$\begin{aligned} M_X(t_1) &= M(t_1, 0) = E[e^{t_1X}] \\ M_Y(t_2) &= M(0, t_2) = E[e^{t_2Y}] \end{aligned}$$

Example 9.1. Suppose $f(x, y) = e^{-y}$, $0 < x < y$. Find $M(t_1, t_2)$ and $M_X(t)$.

$$\begin{aligned} M(t_1, t_2) = E[e^{t_1X+t_2Y}] &= \int_0^\infty \big(\int_0^y e^{t_1x}e^{t_2y}f(x, y)\,\mathrm{d}x\big)\,\mathrm{d}y \\ &= \int_0^\infty e^{(t_2-1)y}\big(\int_0^y e^{t_1x}\,\mathrm{d}x\big)\,\mathrm{d}y \\ &= \frac{1}{t_1}\int_0^\infty e^{(t_2-1)y}e^{t_1y} - 1\,\mathrm{d}y \\ &= \frac{1}{t_1}\int_0^\infty e^{(t_2-1)y}\cdot(e^{t_1y} - 1)\,\mathrm{d}y \\ &= \frac{1}{t_1}\big(\int_0^\infty e^{(t_1+t_2-1)y}\,\mathrm{d}y - \int_0^\infty e^{(t_2-1)y}\,\mathrm{d}y\big) \end{aligned}$$

Note that

$$\int_0^\infty e^{(t_1+t_2-1)y}\,\mathrm{d}y = \frac{1}{t_1+t_2-1}e^{(t_1+t_2-1)y}\,\Big|_0^\infty$$

Taking the limit

$$\lim_{y\to\infty} e^{(t_1+t_2-1)y} < \infty = 0$$

iff $t_1+t_2-1<0$. Similarly $t_2-1<0$ must hold from our other integral.
So we have

$$\begin{aligned} M(t_1,t_2) &= \frac{1}{t_1}\Big(\frac{1}{t_1+t_2-1}(0-1) + \frac{1}{t_2-1}(0-1)\Big) \\ &= \frac{1}{(t_2-1)(t_1+t_2-1)} \end{aligned}$$

For $M_X(t)$ we have

$$M_X(t) = M(t,0) = \frac{1}{1-t}$$

where $t<1$ (from our two constraints on t_1,t_2).
For $M_Y(t)$ we have

$$M_Y(t) = M(0,t) = \frac{1}{(1-t)^2}$$

where $t_2<1$ (from our two constraints on t_1,t_2).
Recall $X \sim GAM(\alpha,\beta)$ has $M_X(t) = \frac{1}{(1-\beta t)^\alpha}$, $t<\frac{1}{\beta}$.
Due to the uniqueness of mgf, $X\sim GAM(1,1)$ and $Y\sim GAM(2,1)$.

9.2 Independence and joint mgfs

X and Y are independent if and only if

$$M(t_1,t_2) = M_X(t_1)M_Y(t_2) \qquad \forall t_1\in B_1, \forall t_2 \in B_2$$

where $M_X(t_1)$ exists in B_1 and $M_Y(t_2)$ exists in B_2 (i.e. the bounds on t_1,t_2 such that $M(t_1,t_2)$ is well-defined).

Example 9.2. From our previous example where

$$M(t_1,t_2) = \frac{1}{(t_2-1)(t_1+t_2-1)}$$

and $M_X(t)=\frac{1}{1-t}$ and $M_Y(t)=\frac{1}{(1-t)^2}$, clearly $M(t_1,t_2)\neq M_X(t_1)M_Y(t_2)$ so X,Y are not independent.

9.3 Summary of methods for verifying independence

The following are equivalent(TFAE) for showing independence of two r.v.s X,Y:

joint pmf/pdf Show $f(x,y)=f_1(x)f_2(y)$

joint cdf Show $F(x,y)=F_1(x)F_2(y)$

Factorization Theorem Show $f(x,y)=h(x)g(y)$ and support set is the rectangular Cartesian product of the individual support sets.

conditional pdf Show $f_1(x\mid y)=f_1(x)$.

joint mgf Show $M(t_1, t_2) = M_X(t_1)M_Y(t_2)$ (for all $(t_1, t_2) \in B$).

10 October 1, 2018

10.1 Expectations/moments from mgf

Suppose X and Y have joint mgf $M(t_1, t_2)$ for all $t_1 \in (-h_1, h_1), t_2 \in (-h_2, h_2)$, some $h_1, h_2 > 0$. Find $E[XY^2]$ and $E[X^kY^j]$, $k, j = 0, 1, 2, \ldots$.

Proof. We can use the moment generating functions to find the expectations.
In the continuous case

$$M(t_1, t_2) = \int \Big(\int e^{t_1 x} e^{t_2 y} f(x, y)\, \mathrm{d}x \Big)\, \mathrm{d}y$$

Thus we have

$$\begin{aligned}
\frac{\partial M(t_1, t_2)}{\partial t_1} &= \frac{\partial}{\partial t_1} \int \Big(\int e^{t_1 x} e^{t_2 y} f(x, y)\, \mathrm{d}x \Big)\, \mathrm{d}y \\
&= \int \Big(\int \big(\frac{\partial}{\partial t_1} e^{t_1 x}\big) e^{t_2 y} f(x, y)\, \mathrm{d}x \Big)\, \mathrm{d}y \\
&= \int \Big(\int x e^{t_1 x} e^{t_2 y} f(x, y)\, \mathrm{d}x \Big)\, \mathrm{d}y
\end{aligned}$$

We then take

$$\begin{aligned}
\frac{\partial^2}{\partial t_1 \partial t_2} M(t_1, t_2) &= \frac{\partial}{\partial t_2} \big(\frac{\partial}{\partial t_1} M(t_1, t_2)\big) \\
&= \frac{\partial}{\partial t_2} \int \Big(\int x e^{t_1 x} e^{t_2 y} f(x, y)\, \mathrm{d}x \Big)\, \mathrm{d}y \\
&= \int \Big(\int x e^{t_1 x} \big(\frac{\partial}{\partial t_2} e^{t_2 y}\big) f(x, y)\, \mathrm{d}x \Big)\, \mathrm{d}y \\
&= \int \Big(\int x e^{t_1 x} y e^{t_2 y} f(x, y)\, \mathrm{d}x \Big)\, \mathrm{d}y
\end{aligned}$$

Once more

$$\begin{aligned}
\frac{\partial^3}{\partial t_1 \partial t_2^2} M(t_1, t_2) &= \frac{\partial}{\partial t_2} \big(\frac{\partial^2}{t_1 t_2} M(t_1, t_2)\big) \\
&= \int \Big(\int x e^{t_1 x} y^2 e^{t_2 y} f(x, y)\, \mathrm{d}x \Big)\, \mathrm{d}y
\end{aligned}$$

Thus if we continue in this fashion

$$\frac{\partial^{k+j}}{\partial t_1^k \partial t_2^j} M(t_1, t_2) = \int \Big(\int x^k e^{t_1 x} y^j e^{t_2 y} f(x, y)\, \mathrm{d}x \Big)\, \mathrm{d}y$$

To find $E[XY^2]$, we simply let $t_1 = t_2 = 0$ in $\frac{\partial^2}{\partial t_1 \partial t_2} M(t_1, t_2)$

$$\begin{aligned}
\Big(\frac{\partial^3}{\partial t_1 \partial t_2^2} M(t_1, t_2)\Big) \Big|_{t_1=t_2=0} &= \int \Big(\int x e^{0x} y^2 e^{0y} f(x, y) \, \mathrm{d}x \Big) \, \mathrm{d}y \\
&= \int \Big(\int x y^2 f(x, y) \, \mathrm{d}x \Big) \, \mathrm{d}y \\
&= E[XY^2]
\end{aligned}$$

Similarly

$$\Big(\frac{\partial^{k+j}}{\partial t_1^k \partial t_2^j} M(t_1, t_2)\Big) \Big|_{t_1=t_2=0} = E[X^k Y^j]$$

This also holds for $E[X^k]$ where

$$\Big(\frac{\partial^k}{\partial t_1^k} M(t_1, t_2)\Big) \Big|_{t_1=t_2=0} = E[X^k]$$

i.e. $j = 0$. $\square$

10.2 Multinomial distribution

Definition 10.1 (Multinomial distribution)**.** Let $(X_1, X_2, \ldots, X_k) \sim MULT(n, p_1, p_2, \ldots, p_k)$ where

$$f(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1! x_2! \ldots x_k! (n - \sum_{i=1}^k x_i)!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} (1 - \sum_{i=1}^k p_i)^{n - \sum_{i=1}^k x_i}$$

where $0 \leq x_i \leq n$, $0 \leq \sum_{i=1}^k x_i \leq n$, $0 \leq p_i \leq 1$, $0 \leq \sum_{i=1}^k p_i \leq 1$.

Remark 10.1. The k random variables represents a random sample of size n where each unit in this random sample could be one of $k + 1$ types with corresponding probabilities $p_1, p_2, \ldots, p_k, 1 - \sum_{i=1}^k p_i$ and x_i is the number elements of the ith type.

Remark 10.2. Binomial $BIN(n, p)$ is a special case of $MULT$ i.e. there are 2 types with probabilities p and $1 - p$ i.e. $MULT(n, p)$ with $k = 1$.

Exercise 10.1 (Hardy-Weinberg law of genetics)**.** We have a random sample of size n from the population. Each unit/person in this sample could be one of 3 genotypes: "AA" with probability $p_1 = \theta^2$, "Aa" with $p_2 = 2\theta(1 - \theta)$, and "aa" with probability $p_3 = (1 - \theta)^2$, $0 < \theta < 1$ i.e. $0 < p_i < 1$ and $\sum_{i=1}^3 p_i = 1$.
Let X_1, X_2 be the number of type "AA" and "Aa", respectively.
Thus

$$P[X_1 = x_1, X_2 = x_2] = \frac{n!}{x_1! x_2! (n - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n - x_1 - x_2}$$

where $0 \leq x_1, x_2 \leq n$ and $0 \leq x_1 + x_2 \leq n$ i.e. $(X_1, X_2) \sim MULT(n, p_1, p_2)$.

10.3 Mgf of multinomial distribution

Note that the MGF for $MULT(n, p_1, p_2)$

$$\begin{aligned} M(t_1, t_2) &= E[e^{t_1 X_1 + t_2 X_2}] \\ &= \sum_{x_1=0}^{n} \sum_{x_2=0}^{n-x_1} e^{t_1 x_1} e^{t_2 x_2} f(x_1, x_2) \\ &= \sum_{x_1=0}^{n} \sum_{x_2=0}^{n-x_1} e^{t_1 x_1} e^{t_2 x_2} \frac{n!}{x_1! x_2! (n - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n - x_1 - x_2} \end{aligned}$$

Recall the Multinomial series identity where

$$(a + b + c)^n = \sum_{x_1=0}^{n} \sum_{x_2=0}^{n-x_1} \frac{n!}{x_1! x_2! (n - x_1 - x_2)!} a^{x_1} b^{x_2} c^{n - x_1 - x_2}$$

for any $a, b, c \in \mathbb{R}$. Thus we have

$$M(t_1, t_2) = (e^{t_1} p_1 + e^{t_2} p_2 + 1 - p_1 - p_2)^n$$

For $e^{t_1} p_1, e^{t_2} p_2 \in \mathbb{R}$, we require $t_1, t_2 \in \mathbb{R}$.
In general for $MULT(n, p_1, \ldots, p_k)$

$$M(t_1, \ldots, t_k) = (e^{t_1} p_1 + \ldots + e^{t_k} p_k + 1 - \sum_{i=1}^{k} p_i)^n$$

10.4 Subset of multinomial is multinomial

Claim. Any "subset" of a multinomial still has a multinomial distribution.
For example suppose we had $(X_1, \ldots, X_6) \sim MULT(n, p_1, \ldots, p_6)$. We have $(X_1, X_3, X_5) \sim MULT(n, p_1, p_3, p_5)$.

Proof. Note that $M(t_1, \ldots, t_6) = (e^{t_1} p_1 + \ldots + e^{t_6} p_6 + 1 - \sum_{i=1}^{k} p_i)^n$, thus

$$\begin{aligned} M_{X_1, X_3, X_5}(t_1, t_3, t_5) &= E[e^{t_1 X_1 + t_3 X_3 + t_5 X_5}] \\ &= E[e^{t_1 X_1 + 0 X_2 + t_3 X_3 + 0 X_4 + t_5 X_5 + 0 X_6}] \\ &= M(t_1, t_2 = 0, t_3, t_4 = 0, t_5, t_6 = 0) \\ &= (e^{t_1} p_1 + e^0 p_2 + e^{t_3} p_3 + e^0 p_4 + e^{t_5} p_5 + e^0 p_6 + 1 - \sum_{i=1}^{6} p_i)^n \\ &= (e^{t_1} p_1 + p_2 + e^{t_3} p_3 + p_4 + e^{t_5} p_5 + p_6 + 1 - \sum_{i=1}^{6} p_i)^n \\ &= (e^{t_1} p_1 + e^{t_3} p_3 + e^{t_5} p_5 + 1 - p_1 - p_3 - p_5)^n \end{aligned}$$

which is the mgf of $MULT(n, p_1, p_3, p_5)$. By the uniqueness of mgfs our claim holds. $\square$

11 October 3, 2018

11.1 More multinomial problems

Example 11.1. Let $T = x_i + x_j$, $1 \le i \le j \le k$.

Claim. We claim $T \sim BIN(n, p_i + p_j)$.

Proof. For $(x_i, x_j) \sim MULT(n, p_i, p_j)$ we have the mgf $M(t_i, t_j) = (e^{t_i}p_i + e^{t_j}p_j) + 1 - p_i - p_j)^n$ for all $t_i, t_j \in \mathbb{R}$. The mgf of T is $M_T(t) = E[e^{tT}] = E[e^{t(X_i+X_j)}] = E[e^{tX_i+tX_j}]$.
Thus

$$\begin{aligned}M_T(t) &= M(t_i = t, t_j = t)\\ &= (e^t p_i + e^t p_j + 1 - p_i - p_j)^n\\ &= (e^t(p_i + p_j) + 1 - (p_i + p_j))^n\end{aligned}$$

Recall the mgf of $X \sim BIN(n, p)$ is $M_X(t) = (e^t p + 1 - p)^n$ for all $t \in \mathbb{R}$, thus $T = x_i + x_j \sim BIN(n, p_i + p_j)$ by uniqueness of mgf. □

Claim. We claim $Cov(X_i, X_j) = -np_ip_j$.

Proof. Note that $Cov(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j)$. Also

$$E(X_iX_j) = \big(\frac{\partial^2}{\partial t_i \partial t_j} M(t_i, t_j)\big) \mid_{t_i=t_j=0}$$

We have

$$\begin{aligned}\frac{\partial}{\partial t_i} M(t_i, t_j) &= \frac{\partial}{\partial t_i}(e^{t_i}p_i + e^{t_j}p_j + 1 - p_i - p_j)^n\\ &= n(e^{t_i}p_i + e^{t_j}p_j + 1 - p_i - p_j)^{n-1} \cdot e^{t_i}p_i\end{aligned}$$

Furthermore

$$\begin{aligned}\frac{\partial^2}{\partial t_i \partial t_j} M(t_i, t_j) &= \frac{\partial}{\partial t_j}(\frac{\partial}{\partial t_i} M(t_i, t_j))\\ &= n(n-1)(e^{t_i}p_i + e^{t_j}p_j + 1 - p_i - p_j)^{n-2} \cdot e^{t_i}p_i \cdot e^{t_j}p_j\end{aligned}$$

Therefore

$$\begin{aligned}E(X_iX_j) &= n(n-1)(e^0p_i + e^0p_j + 1 - p_i - p_j)^{n-2}e^0p_ie^0p_j\\ &= n(n-1)p_ip_j\end{aligned}$$

So we have

$$Cov(X_i, X_j) = n(n-1)p_ip_j - (np_i)(np_j) = -np_ip_j$$

as $X_i \sim BIN(n, p_i)$ and $E(X_i) = np_i$. □

Claim. We claim $(X_i \mid X_j = x_j) \sim BIN(n - x_j, \frac{p_i}{1-p_j})$.

Note that $(X_i, X_j) \sim MULT(n, p_i, p_j)$ and $X_j \sim BIN(n, p_j)$, thus

$$
\begin{aligned}
f(x_i \mid x_j) &= \frac{f(x_i, x_j)}{f(x_j)} \\
&= \frac{\frac{n!}{x_i! x_j! (n-x_i-x_j)!} p_i^{x_i} p_j^{x_j} (1-p_i-p_j)^{n-x_i-x_j}}{\frac{n!}{(n-x_j)! x_j!} p_j^{x_j} (1-p_j)^{n-x_j}} \\
&= \frac{(n-x_j)!}{(n-x_j-x_i)! x_i!} \frac{p_i^{x_i} (1-p_i-p_j)^{n-x_i-x_j}}{(1-p_j)^{n-x_j}} \\
&= \frac{(n-x_j)!}{(n-x_j-x_i)! x_i!} \frac{p_i^{x_i}}{(1-p_j)^{x_i}} \frac{(1-p_i-p_j)^{n-x_i-x_j}}{(1-p_j)^{n-x_i-x_j}} \\
&= \frac{(n-x_j)!}{(n-x_j-x_i)! x_i!} \big(\frac{p_i}{(1-p_j)}\big)^{x_i} \big(\frac{(1-p_i-p_j)}{(1-p_j)}\big)^{n-x_i-x_j} \\
&= \frac{(n-x_j)!}{(n-x_j-x_i)! x_i!} \big(\frac{p_i}{(1-p_j)}\big)^{x_i} \big(1-\frac{p_i}{(1-p_j)}\big)^{n-x_i-x_j}
\end{aligned}
$$

i.e. $f_i(x_i \mid x_j)$ is the same as the pmf of $BIN(n-x_j, \frac{p_i}{1-p_j})$ so the claim holds.

Claim. We claim $X_i \mid X_i + X_j = t \sim BIN(t, \frac{p_i}{p_i+p_j})$.

Proof. Note that

$$P(X_i = x_i, X_i + X_j = t) = P(X_i = x_i, X_j = t - x_i)$$

which is just our joint pmf.
Also from before we have $P(X_i + X_j = t)$ is the pmf of $T = X_i + X_j$ and $T \sim BIN(n, p_i + p_j)$.
Thus we have

$$
\begin{aligned}
f_i(x_i \mid t) &= \frac{P(X_i = x_i, X_j = t - x_i)}{P(T = t)} \\
&= \frac{\frac{n!}{x_i!(t-x_i)!(n-t)!} p_i^{x_i} p_j^{t-x_i} (1-p_i-p_j)^{n-t}}{\frac{n!}{t!(n-t)!} (p_i+p_j)^t (1-p_i-p_j)^{n-t}} \\
& \frac{t!}{x_i!(t-x_i)} \frac{p_i^{x_i} p_j^{t-x_i}}{(p_i+p_j)^t} \\
& \frac{t!}{x_i!(t-x_i)} \big(\frac{p_i}{p_i+p_j}\big)^{x_i} \big(1-\frac{p_i}{p_i+p_j}\big)^{t-x_i}
\end{aligned}
$$

which is the pmf of $BIN(t, \frac{p_i}{p_i+p_j})$. □

11.2 Bivariate normal distribution

Recall for a univariate normal distribution $X \sim N(\mu, \sigma^2)$

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{(2\pi)^{\frac{1}{2}} (\sigma^2)^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)}
\end{aligned}
$$

The bivariate normal distribution for $\vec{x} = (x_1, x_2)^T$ is denoted as $X \sim BVN(\vec{\mu}, \Sigma)$ where $\vec{\mu} = (E(X_1), E(X_2))^T$ and

$$\Sigma = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{bmatrix}$$

Notice that $Cov(X_1, X_2) = Cov(X_2, X_1)$ i.e. Σ is symmetric and positive definite.
We define the pdf for the bivariate normal distribution as

$$f(x_1, x_2) = \frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})}$$

for all $x_1, x_2 \in \mathbb{R}$.

12 October 5, 2018

12.1 Remarks of bivariate normal

Remark 12.1. 1.

$$\vec{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} E[X_1] \\ E[X_2] \end{bmatrix}$$

2.

$$\Sigma = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & P\sigma_1\sigma_2 \\ P\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

where

$$P = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)}\sqrt{Var(X_2)}}$$

and $-1 < P < 1$ (note if $P = \pm 1$ then Σ is not full rank thus Σ^{-1} does not exist).

3.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2 = 1$$

(useful for finding $M(t_1, t_2)$).

4. Σ is positive definite (symmetric by definition) i.e. for all

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

we have $y^T \Sigma y > 0$.

That is: both eigenvalues of Σ are positive.

Remark 12.2. If $X \sim BVN(\mu + \Sigma t, \Sigma)$ where $t = [t_1, t_2]^T$ then the joint pdf

$$g(x_1, x_2) = \frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu-\Sigma t)^T \Sigma^{-1} (x-\mu-\Sigma t)}$$

for $x \in \mathbb{R}^2$ then

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2 = 1$$

12.2 Mgf of bivariate normal

Claim. For $BVN(\mu, \Sigma)$ we claim its mgf is

$$M(t_1, t_2) = e^{t^T\mu + \frac{1}{2}t^T\Sigma t}$$

where $t = [t_1, t_2]^T \in \mathbb{R}^2$.

Proof. Note that the following properties hold in general

1. $a^T b = b^T a$ for $a, b \in \mathbb{R}^2$.
2. $\Sigma = \Sigma^T$ since Σ is symmetric.
3. $\Sigma\Sigma^{-1} = \Sigma^{-1}\Sigma = I_{2\times 2}$
4. $Ia = a$ and $a^T I = a^T$.
5. $(\Sigma t)^T = t^T \Sigma^T$

We have

$$M(t_1, t_2) = E[e^{t^T x}]$$

where $t = [t_1, t_2]^T, x = [x_1, x_2]^T$.

$$\begin{aligned} M(t_1, t_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t^T x}\frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}\left((x-\mu)^T\Sigma^{-1}(x-\mu)-2t^T x\right)}\,\mathrm{d}x_1\,\mathrm{d}x_2 \end{aligned}$$

If we can show that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)-2t^T x}\,\mathrm{d}x_1\,\mathrm{d}x_2 = e^{t^T\mu+\frac{1}{2}t^T\Sigma t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu^*)^T\Sigma^{-1}(x-\mu^*)}\,\mathrm{d}x_1\,\mathrm{d}x_2$$

for some μ^* then we are done since the integral on the right is just the total probability of a bivariate r.v. $N(\mu^*, \Sigma)$ which is 1.

$$\begin{aligned} &e^{t^T\mu+\frac{1}{2}t^T\Sigma t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu^*)^T\Sigma^{-1}(x-\mu^*)}\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}\left((x-\mu^*)^T\Sigma^{-1}(x-\mu^*)-2(t^T\mu+\frac{1}{2}t^T\Sigma t)\right)}\,\mathrm{d}x_1\,\mathrm{d}x_2 \end{aligned}$$

We notice that every term is the same except for the exponential terms: thus we need to show the exponential terms are equivalent, specifically the terms in the exponent:

$$(x-\mu)^T\Sigma^{-1}(x-\mu) - 2t^T x = (x-\mu^*)^T\Sigma^{-1}(x-\mu^*) - 2(t^T\mu + \frac{1}{2}t^T\Sigma t)$$

We claim $\mu^* = \mu + \Sigma t$. We have on the RHS

$$\begin{aligned}
&(x-\mu-\Sigma t)^T\Sigma^{-1}(x-\mu-\Sigma t) - 2(t^T\mu + \frac{1}{2}t^T\Sigma t)\\
=&\big((x-\mu)^T - (\Sigma t)^T\big)\Sigma^{-1}\big((x-\mu)-(\Sigma t)\big) - 2(t^T\mu + \frac{1}{2}t^T\Sigma t)\\
=&(x-\mu)^T\Sigma^{-1}(x-\mu) + (\Sigma t)^T\Sigma^{-1}(\Sigma t) - (\Sigma t)^T\Sigma^{-1}(x-\mu) - (x-\mu)^T\Sigma^{-1}(\Sigma t) - 2(t^T\mu + \frac{1}{2}t^T\Sigma t)\\
=&(x-\mu)^T\Sigma^{-1}(x-\mu) + t^T\Sigma t - t^T(x-\mu) - (x-\mu)^T t - 2(t^T\mu + \frac{1}{2}t^T\Sigma t)\\
=&(x-\mu)^T\Sigma^{-1}(x-\mu) - 2t^T(x-\mu) - 2(t^T\mu + \frac{1}{2}t^T\Sigma t)\\
=&(x-\mu)^T\Sigma^{-1}(x-\mu) - 2t^T x
\end{aligned}$$

as desired. □

12.3 Joint cdf from pdf

Example 12.1. Let $f(x,y) = 2e^{-x}e^{-y}$ for $0 < x < y$.
We want to find $F(x,y) = P(X \le x, Y \le y)$.
If $x \le 0$ or $y \le 0$ then $F(x,y) = 0$ (does not intersect our support set).
When $0 < x \le y$ (note $P(x = y) = 0$ so it does not matter/influence or cdf), we have

$$\int_0^x \int_x^y f(x,y)\,\mathrm{d}y\,\mathrm{d}x = (1-e^{-2x}) - 2e^{-y}(1-e^{-x})$$

For $0 < y < x$, we note that since $x < y$ in our support set we are really calculating $F(y,y)$ so from above

$$F(y,y) = (1-e^{-2y}) - 2e^{-y}(1-e^{-y}) = 1 + e^{-2y} - 2e^{-y}$$

If we want to find $F_1(x) = \lim_{y\to\infty} F(x,y)$.
Note for the region $x \le 0$, we have $F_1(x) = 0$.
For the region $0 < x < \infty$, we take our $F(x,y)$ and take the limit thus we get $F_1(x) = 1 - e^{-2x}$.

13 October 12, 2018

13.1 Marginal pdf of bivariate normal

Exercise 13.1. $X \sim N(\mu, \sigma^2)$ then $M_X(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2}$ for $t \in \mathbb{R}$.

Claim. If $X \sim BVN(\mu, \Sigma)$ then $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$.

Proof. Note that

$$\begin{aligned}
M_X(t_1,t_2) &= \exp(t^T\mu + \frac{1}{2}t^T\Sigma t) \\
&= \exp((t_1,t_2)(\mu_1,\mu_2)^T + \frac{1}{2}(t_1,t_2)\begin{bmatrix}\sigma_1^2 & P\sigma_1\sigma_2 \\ P\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}(t_1,t_2)^T) \\
&= \exp(t_1\mu_1 + t_2\mu_2 + \frac{1}{2}t_1^2\sigma_1^2 + \frac{1}{2}t_2^2\sigma_2^2 + t_1t_2P\sigma_1\sigma_2) \\
&= \exp(t_1\mu_1 + \frac{1}{2}t_1^2\sigma_1^2)\exp(t_2\mu_2 + \frac{1}{2}t_2^2\sigma_2^2)\exp(t_1t_2P\sigma_1\sigma_2)
\end{aligned}$$

Thus

$$\begin{aligned}
M_{X_1}(t_1) = M_X(t_1,0) &= \exp(t_1\mu_1 + \frac{1}{2}t_1^2\sigma_1^2)\exp(0)\exp(0) \\
&= \exp(t_1\mu_1 + \frac{1}{2}t_1^2\sigma_1^2)
\end{aligned}$$

which is the mgf of $N(\mu_2,\sigma_2^2)$. Similarly $M_{X_2}(t_2)$ has the same mgf as $N(\mu_2,\sigma_2^2)$. □

Claim. X_1 and X_2 are independent iff $P=0$.

Proof. Recall X_1,X_2 are independent iff $M_X(t_1,t_2)=M_{X_1}(t_1)M_{X_2}(t_2)$ for all $t_1,t_2\in\mathbb{R}$.
Thus the LHS is $M(t_1,t_2)=M_{X_1}(t_1)M_{X_2}(t_2)\exp(t_1t_2P\sigma_1\sigma_2)$ and the RHS is $M_{X_1}(t_1)M_{X_2}(t_2)$ therefore LHS = RHS iff $\exp(t_1t_2P\sigma_1\sigma_2)=1$ i.e. $t_1t_2P\sigma_1\sigma_2=0$ for all $t_1,t_2\in\mathbb{R}$, thus $P=0$ since $\sigma_1,\sigma_2>0$. □

Claim. If $Y=c^TX\sim N(c^T\mu, c^T\Sigma c)$ where $c=(c_1,c_2)^T\neq(0,0)^T$.

Proof. Note that $M_Y(t)=E[\exp(tY)]=E[\exp(tc^TX)]=E[\exp((tc)^TX)]$. Let $t^*=tc$ then we have $M_Y(t)=E[\exp((t^*)^TX)]=M(t_1^*,t_2^*)=\exp(t^*\mu+\frac{1}{2}(t^*)^T\Sigma t^*)$.
Thus we have

$$\begin{aligned}
M_Y(t) &= \exp((tc)^T\mu + \frac{1}{2}(tc)^T\Sigma(tc)) \\
&= \exp(t(c^T\mu) + \frac{1}{2}t^2(c^T\Sigma c)
\end{aligned}$$

which is the mgf of $N(c^T\mu, c^T\Sigma c)$ due to the uniqueness theorem of mgf. □

Claim. Let $Y=AX+b$ where $A\in\mathbb{R}^{2\times 2}$ and $b=(b_1,b_2)^T$.
Then $Y\sim BVN(A\mu+b, A\Sigma A^T)$.

Proof. Exercise (similar to proof above). □

14 October 15, 2018

14.1 Bivariate transformation

Suppose we wanted to transform a bivariate r.v. $(X,Y)\to(U,V)$ or to U only. We can then find the distribution of (U,V) (or U only) based on (X,Y).
There are two methods that are analogous to the ones for univariate random variables

Method 1 1-to-1 transformation $(X, Y) \iff (U, V)$

Method 2 cdf technique

Recall in the univariate case with the 1-to-1 technique:

Example 14.1. Let $f(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}}$ for $x > \alpha$ $(\alpha, \beta > 0)$.
Find the pdf of $Y = \beta \log(\frac{X}{\alpha})$.
Note that $Y = h(X) = \beta \log(\frac{X}{\alpha})$ is a 1-to-1 function as $\alpha, \beta > 0$ and $\log(\cdot)$ is monotonically increasing.
Thus $X = \alpha e^{\frac{Y}{\beta}} = h^{-1}(Y)$. Note that

$$\frac{d}{dy}h^{-1}(y) = \frac{\alpha}{\beta}e^{\frac{y}{\beta}}$$

Thus we have

$$g(y) = f(h^{-1}(y))\left|\frac{d}{dy}h^{-1}(y)\right| = \frac{\beta\alpha^\beta}{(\alpha e^{\frac{y}{\beta}})^{\beta+1}}\frac{\alpha}{\beta}e^{\frac{y}{\beta}} = e^{-y}$$

Note that the support set of Y is $y > 0$.

Similarly for this univariate case

Example 14.2. Let $X \sim EXP(1)$, $X > 0$. Let $Z = X^2$. Find the pdf of Z.

Method 1 $Z = h(X) = X^2$ is 1-to-1 since X is positive. Therefore $X = +\sqrt{Z} = h^{-1}(Z)$ thus $\frac{d}{dz} = \frac{1}{2}z^{-0.5}$ and so

$$g(z) = e^{-z^{0.5}}\frac{1}{2}z^{-0.5} \qquad z > 0$$

Method 2 Note that $G(z) = P(Z \leq z)$, which is 0 if $z \leq 0$.

For $z > 0$ we have

$$\begin{aligned} P(Z \leq z) = P(X^2 \leq z) &= P(X \leq \sqrt{z}) && X \text{ is positive} \\ &= F(\sqrt{z}) \\ &= 1 - e^{-\sqrt{z}} \end{aligned}$$

where $F(x) = 1 - e^{-x}$ if $x > 0$ and 0 if $x \leq 0$.

Thus

$$g(z) = \frac{d}{dz}G(z) = f(\sqrt{z})\frac{1}{2}z^{-0.5} = e^{-\sqrt{z}}\frac{1}{2}z^{-0.5}$$

where $z > 0$.

What about a bivariate case?

Example 14.3. Recall that with $f(x, y) = ke^{-x}e^{-y}$ for $0 < x < y$ we wanted to find $P(X + Y \geq 1)$, i.e. if $U = X + Y$ we are finding $P(U \geq 1)$. However $(X, Y) \to U$ is not a 1-to-1 transformation: one pair (X, Y) corresponds to one U but one U does not correspond to a unique (X, Y), thus we cannot transform our bivariate to a univariate distribution.

Example 14.4. Let $f(x, y) = 3y$ for $0 < x < y < 1$. Find the pdf of $U = XY$ i.e. we want to map $(X, Y) \to U$.

Note that

$$\begin{aligned}G(u) &= P(U \le u)\\ &= P(XY \le u)\\ &= \begin{cases} 0 & \text{if } u \le 0\\ (*) & \text{if } 0 < u < 1\\ 1 & \text{if } u \ge 1\end{cases}\end{aligned}$$

where the above follows since the support set is $0 < x < y < 1 \Rightarrow 0 < x, y < 1$.
We have $(*)$ as $P(xY \le u) = P(Y \le \frac{u}{x})$ for $0 < u < 1$.

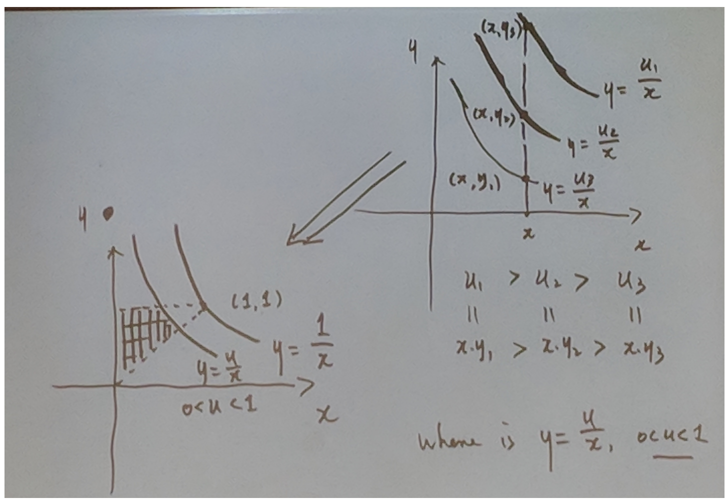


Figure 14.1: Graphs of $y = \frac{u_i}{x}$ for various $0 < u_i < 1$ (top right); graph of $y = \frac{1}{x}$, $y = \frac{u}{x}$ for $0 < u < 1$ and the support triangle region (bottom left). We want to integrate over the shaded area to find $P(Y \le \frac{u}{x})$.

For a given fixed x, if we look at the $y = \frac{u}{x}$ for $0 < u < 1$, we see that for $P(Y \le \frac{u}{x})$ we are essentially integrating the area underneath $y = \frac{u}{x}$ that intersects with our support which is the region in the unit square of the first quadrant above $y = x$ since $y > x$.

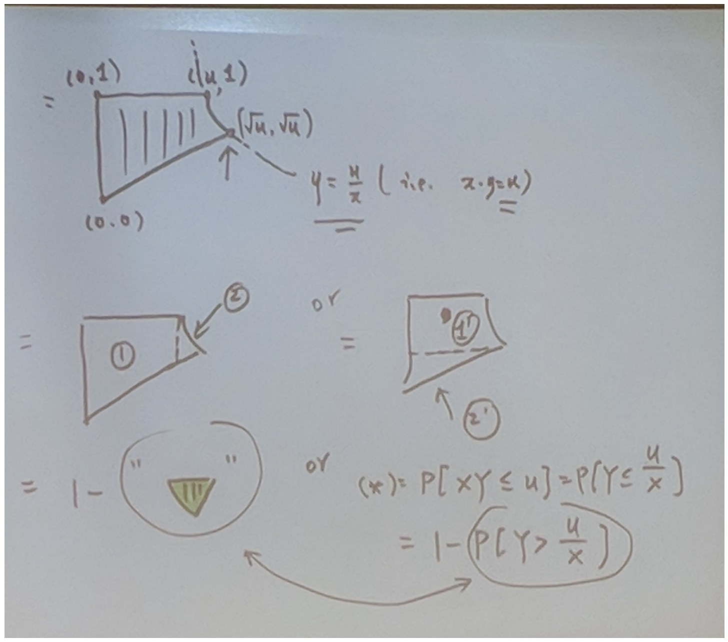


Figure 14.2: Area of integration where we can either integrate over region (1) and (2), (1') and (2'), or integrate over the green triangle and subtract it from 1.

Note the limits of integration in the image

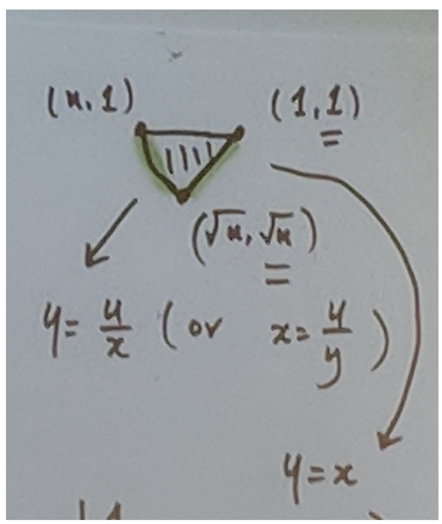


Figure 14.3: Bounds for integrations for the green shaded we are integrating over.

So we'd like to find

$$\begin{aligned}
P(Y \le \frac{u}{x}) &= 1 - P(Y > \frac{u}{x}) \\
&= 1 - \int_{\sqrt{u}}^{1} \big(\int_{\frac{u}{y}}^{y} f(x,y)\,\mathrm{d}x \big)\,\mathrm{d}y \\
&= 1 - \int_{\sqrt{u}}^{1} \big(\int_{\frac{u}{y}}^{y} 3y\,\mathrm{d}x \big)\,\mathrm{d}y \\
&= 1 - \int_{\sqrt{u}}^{1} 3y(y - \frac{u}{y})\,\mathrm{d}y \\
&= 1 - (y^3\Big|_{\sqrt{u}}^{1} - 3u(1-\sqrt{u})) \\
&= 3u - 2u\sqrt{u} \qquad 0 < u < 1
\end{aligned}$$

Thus we have

$$G(u) = \begin{cases} 0 & \text{if } u \le 0 \\ 3u - 2u\sqrt{u} & \text{if } 0 < u < 1 \\ 1 & \text{if } u \ge 1 \end{cases}$$

so $g(u) = \frac{d}{du}G(u) = 3 - 3\sqrt{u}$ where $G(u)$ is differentiable for $0 < u < 1$.
$G(u) = 0$ if $u \le 0$ implies $g(u) = 0$ if $u < 0$, similarly if $u \ge 1$ we have $g(u) > 1$.
What happens if $u = 0$ or $u = 1$? Note that at $u = 0$ $G(u)$ is NOT differentiable. Question: Is $G(u)$ differentiable at $u = 1$?
Regardless, our answer is $g(u) = 3 - 3\sqrt{u}$ for $0 < u < 1$.

Exercise 14.1. Find the pdf of $V = \frac{Y}{X}$.

Example 14.5. Let X_i be iid with common pdf and cdf $f(x)$ and $F(x)$, respectively, $i = 1, \dots, n$.
Find the pdf of $S = \max(X_1, \dots, X_n)$ and $T = \min(X_1, \dots, X_n)$ separately.
Notice that $(x_1, \dots, x_n) \to S$ (or T) is NOT 1-1.
So

$$\begin{aligned}
G(s) = P(S \le s) &= P(max(X_1, \dots, X_n) \le s) \\
&= P(X_1 \le s, X_2 \le s, \dots, X_n \le s) \\
&= P(X_1 \le s) \cdot P(X_2 \le s) \cdot \ldots \cdot P(X_n \le s) && \text{independence} \\
&= F(s)^n
\end{aligned}$$

Thus

$$g(s) = \frac{d}{ds}G(s) = nF(s)^{n-1}f(s)$$

where $F(s)$ and $f(s)$ are the cdf and pdf of X evaluated at s, respectively.

Exercise 14.2. Find the pdf of T.

For 1-to-1 bivariate transformation where we have $(X,Y) \iff (U,V)$, we can find the joint pdf of (U,V) based on the joint pdf of (X,Y).

Recall for $Y = h(X)$ where h is 1-to-1, we have $X = h^{-1}(Y)$ and

$$g(y) = f(h^{-1}(y))\left|\frac{d}{dy}h^{-1}(y)\right|$$

For $U = h_1(X,Y)$ and $V = h_2(X,Y)$ where h_1, h_2 are 1-to-1, then $X = w_1(U,V)$ and $Y = w_2(U,V)$ where w_1, w_2 are the inverses of h_1, h_2, respectively.
Thus

$$g(u,v) = f(w_1, w_2)\left|\frac{\partial(w_1,w_2)}{\partial(u,v)}\right|$$

where

$$\frac{\partial(w_1,w_2)}{\partial(u,v)} = \det\begin{bmatrix}\frac{\partial w_1}{\partial u} & \frac{\partial w_1}{\partial v}\\ \frac{\partial w_2}{\partial u} & \frac{\partial w_2}{\partial v}\end{bmatrix} = \frac{\partial w_1}{\partial u}\cdot\frac{\partial w_2}{\partial v} - \frac{\partial w_1}{\partial v}\cdot\frac{\partial w_2}{\partial u}$$

Let R_{XY} and R_{UV} be the support set of (X,Y) of (U,V), respectively.
Notice that R_{UV} is based on R_{XY} through the bivariate transformation.
Thus the steps for the 1-to-1 technique for bivariate transformations are

1. Verify 1-to-1 transformation
2. Find (w_1, w_2) and $\frac{\partial(w_1,w_2)}{\partial(u,v)}$
3. Find $g(u,v)$
4. Find R_{UV}

15 October 17, 2018

15.1 Verifying 1-to-1 for bivariate functions

The **inverse mapping/function theorem** states that $U = h_1(X,Y)$ and $V = h_2(X,Y)$ are 1-1 if

1. $\frac{\partial h_1}{\partial x}, \frac{\partial h_1}{\partial y}, \frac{\partial h_2}{\partial x}, \frac{\partial h_2}{\partial y}$ are continuous functions of x and y in R_{XY}.
2. $$\frac{\partial(w_1,w_2)}{\partial(u,v)} = \det\begin{bmatrix}\frac{\partial w_1}{\partial u} & \frac{\partial w_1}{\partial v}\\ \frac{\partial w_2}{\partial u} & \frac{\partial w_2}{\partial v}\end{bmatrix} \neq 0$$
   in R_{XY}.

15.2 Examples of bivariate transformations

Example 15.1. Let $X \sim GAM(a,1)$ independent of $Y \sim GAM(b,1)$.
Let $U = X + Y = h_1(X,Y)$, $V = \frac{X}{X+Y} = h_2(X,Y)$.
Find the joint pdf of (U,V): $g(u,v)$.

Solution. We can do this in 4 steps:

Step 1: find $f(x,y)$ Note that $f(x,y) = f_1(x)f_2(y)$ by independence so

$$\frac{1}{\Gamma(a)}x^{a-1}e^{-x}\cdot\frac{1}{\Gamma(b)}y^{b-1}e^{-y}$$

Furthermore

$$R_{XY} = R_X \times R_Y = (0,\infty)\times(0,\infty) = \{(x,y) \mid x,y > 0\}$$

Step 2: Verify 1-to-1 Verify 1-to-1 of h_1, h_2 by inverse mapping theorem

1. Note $\frac{\partial h_1}{\partial x} = 1$, $\frac{\partial h_1}{\partial y} = 1$. Also

$$\frac{\partial h_2}{\partial x} = \frac{(x+y)-x}{(x+y)^2} = \frac{y}{(x+y)^2}$$
$$\frac{\partial h_2}{\partial y} = \frac{-x}{(x+y)^2}$$

   Note both are continuous on R_{XY} (no discontinuity on R_{XY} since $x+y \neq 0$ and they are quotients of continuous functions which is continuous).

2.

$$\begin{aligned}\frac{\partial(h_1,h_2)}{\partial(x,y)} &= \det \begin{bmatrix} \frac{\partial h_1}{\partial x} & \frac{\partial h_1}{\partial y} \\ \frac{\partial h_2}{\partial x} & \frac{\partial h_2}{\partial y} \end{bmatrix} \\ &= \frac{\partial h_1}{\partial x} \cdot \frac{\partial h_2}{\partial y} - \frac{\partial h_1}{\partial y} \cdot \frac{\partial h_2}{\partial x} \\ &= 1\frac{-x}{(x+y)^2} - 1\frac{y}{(x+y)^2} \\ &= \frac{-1}{x+y} \\ &\neq 0 \qquad (x,y) \in R_{XY}\end{aligned}$$

Therefore our functions are indeed 1-to-1 by the inverse mapping theorem.

Step 3: find inverse $g(u,v)$ We find our inverse transformations and $\frac{\partial(w_1,w_2)}{\partial(u,v)}$. Note that we can see

$$X = UV = w_1(U,V)$$
$$Y = U - UV = U(1-V) = w_2(U,V)$$

We also have

$$\begin{aligned}\frac{\partial(w_1,w_2)}{\partial(u,v)} &= \left(\frac{\partial w_1}{\partial u}\right)\left(\frac{\partial w_2}{\partial v}\right) - \left(\frac{\partial w_1}{\partial v}\right)\left(\frac{\partial w_2}{\partial u}\right) \\ &= (v)(-u) - (1-v)(u) \\ &= -u\end{aligned}$$

So we have (where $f(x,y) = \frac{1}{\Gamma(b)\Gamma(a)} x^{a-1}e^{-x}y^{b-1}e^{-y}$)

$$\begin{aligned}g(u,v) &= f(w_1,w_2)\left|\frac{\partial(w_1,w_2)}{\partial(u,v)}\right| \\ &= \frac{1}{\Gamma(b)\Gamma(a)}(uv)^{a-1}e^{-uv}(u(1-v))^{b-1}e^{-u(1-v)} \cdot |-u| \\ &= \frac{1}{\Gamma(b)\Gamma(a)}u^{a+b-1}e^{-u}v^{a-1}(1-v)^{b-1}\end{aligned}$$

Remark 15.1. We can factorize $g(u,v) = f_1(u)f_2(v)$ into a function of u and a function of v.

Step 4: find R_{UV} support We derive this from R_{XY}: note that $R_{XY} = \{(x, y) \mid x, y > 0\}$, where $X = UV$ and $Y = U - UV$, thus we have

$$R_{UV} = \{(u, v) \mid w_1(u, v) = uv > 0, w_2(u, v) = u - uv > 0\}$$

Since $uv > 0$, then $u, v > 0$.

Secondly since $u - uv > 0$, then $u > uv > 0$ so $u, v > 0$ and $v < 1$.

Thus we have $R_{UV} = \{(u, v) \mid u > 0, 0 < v < 1\}$.

That is $R_{UV} = (0, \infty) \times (0, 1)$ is recntagular, so U, V are independent by the remark above and the factorization theorem.

Optional step: Marginal pdfs We claim $U \sim GAM(a + b, 1)$ and $V \sim BETA(a, b)$. Note that

$$g_1(u) = \int_0^1 g(u, v)\,\mathrm{d}v = \frac{1}{\Gamma(a + b)} u^{a+b-1} e^{-u}$$

which is the pdf of $GAM(a + b, 1)$.

Similarly

$$g_2(v) = \int_0^\infty g(u, v)\,\mathrm{d}u = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} v^{a-1} (1 - v)^{b-1}$$

which is the pdf of $BETA(a, b)$.

15.3 Computing determinant from the inverse's determinant

Note that in the previous example, we ca compute $\frac{\partial(w_1, w_2)}{\partial(u, v)}$ indirectly by

$$\frac{\partial(w_1, w_2)}{\partial(u, v)} = [\frac{\partial(h_1, h_2)}{\partial(x, y)}]^{-1}\Big|_{x=w_1(u,v), y=w_2(u,v)}$$

where we need to substitute x and y for $w_1(u, v)$ and $w_2(u, v)$ after computing the inverse of the determinant. Recall from our previous example that we had

$$\frac{\partial(h_1, h_2)}{\partial(x, y)} = \frac{-1}{x + y}$$

thus we have

$$\frac{\partial(w_1, w_2)}{\partial(u, v)} = (\frac{-1}{x + y})^{-1} = -(x + y) = -(uv + u - uv) = -u$$

which agrees with our previous result.

15.4 Bivariate transformations of non 1-to-1 functions

Example 15.2. For $f(x, y) = 3y$ where $0 < x < y < 1$, let $U = XY$. Suppose we wanted to find the pdf of U. Note that $(X, Y) \to U$ is not 1-to-1: we have multiple $f(x, y)$ for the same u. For example if we had $u = 16$, we can either $(x, y) = (2, 8)$ or $(x, y) = (4, 4)$ which maps to different $f(x, y)$ values.
We can include some random variable V to ensure $(X, Y) \leftrightarrow (U, V)$ is 1-to-1, then we can compute the marginal pdf of U via $g_1(u) = \int g(u, v)\,\mathrm{d}v$.
What V do we choose? We claim V is not unique. Let $V = X$, so $U = h_1(X, Y) = XY$ and $V = h_2(X, Y) = X$.

We note that for $u = 16, v = 2$, we only have one $(x, y) = (2, 8)$ that maps to one unique $f(x, y) = 24$ value. Similarly $V = Y$ works as well.

16 October 19, 2018

16.1 Bivariate transformation with dummy second variable

For $U = XY$, find $f(u, v)$ and $f_1(u)$ for some V using the 1-to-1 technique.

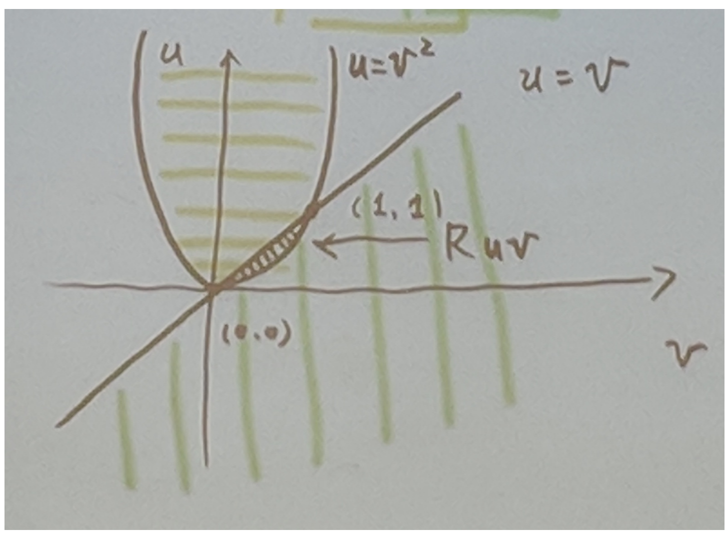
Solution. Step 1: verify 1-to-1 Use inverse mapping theorem and verify the partial derivatives are continuous on R_{XY}.

Step 2: inverse and determinant of Jacoby We let $V = X$. Thus $X = V = w_1(U, V)$ and $Y = \frac{U}{V} = w_2(U, V)$.

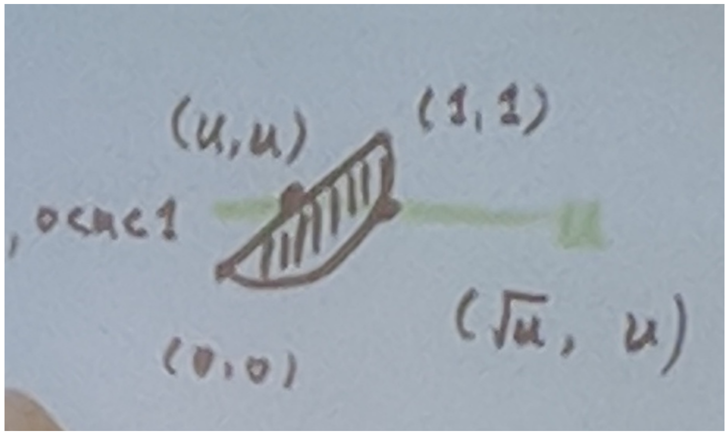
We find that $\frac{\partial(w_1,w_2)}{\partial(u,v)} = \frac{-1}{v}$ (also $\frac{\partial(w_1,w_2)}{\partial(u,v)} = \left(\frac{\partial(h_1,h_2)}{\partial(x,y)}\right)^{-1}\Big|_{x=w_1(u,v),y=w_2(u,v)}$ from step 1).

Step 3: find $g(u, v)$ We find that $g(u, v) = \frac{3u}{v}$.

Step 4: find R_{UV} Recall $R_{XY} = \{(x, y) \mid 0 < x < y < 1\}$, thus $R_{UV} = \{(u, v) \mid 0 < v < \frac{u}{v} < 1\}$. Firstly: $0 < u < 1$ since $u = xy$ and $0 < x < y < 1$ and $0 < v < 1$ since $v = x$. Secondly, $0 < v^2 < u < v$ where u is domain corresponding to the area above $u = v^2$ (parabola) and below $u = v$ (line).



Step 5: find $g_1(u)$ Note that our integration over v bounds for u is $u < v < \sqrt{u}$.



$$
\begin{aligned}
g_1(u) &= \int_u^{\sqrt{u}} g(u,v)\,\mathrm{d}v \\
&= \int_u^{\sqrt{u}} \frac{3u}{v^2}\,\mathrm{d}v \\
&= 3 - 3\sqrt{u} \qquad 0 < u < 1
\end{aligned}
$$

which is the safe pdf of u when we used the cdf technique.

16.2 Box-Mueller transformation example

Let $X, Y \sim UNIF(0,1)$ be iid.
Let

$$
\begin{aligned}
U = h_1(X,Y) &= (-2\log X)^{\frac{1}{2}}\cos(2\pi Y) \\
V = h_2(X,Y) &= (-2\log X)^{\frac{1}{2}}\sin(2\pi Y)
\end{aligned}
$$

Find $g(u,v)$ and marginal distribution of XS and Y.

Solution. Note that our joint pdf for $f(x,y)$ is $f(x,y) = f_1(x)f_2(y) = 1$ where $R_{XY} = \{(x,y) \mid 0 < x, y < 1\} = R_X \times R_Y$ due to independence.
We thus have

$$
g(u,v) = f(w_1,w_2)\left|\frac{\partial(w_1,w_2)}{\partial(u,v)}\right| = \left|\frac{\partial(w_1,w_2)}{\partial(u,v)}\right| = \left|(\frac{\partial(h_1,h_2)}{\partial(x,y)})^{-1}\right|
$$

where

$$
\frac{\partial(h_1,h_2)}{\partial(x,y)} = \frac{\partial h_1}{\partial x}\frac{\partial h_2}{\partial y} - \frac{\partial h_1}{\partial y}\frac{\partial h_2}{\partial x}
$$

Step 1: Verify 1-to-1 Note that

$$\begin{aligned}
\frac{\partial h_1}{\partial x} &= \big(\frac{1}{2}\big)\big(\frac{-2}{x}\big)(-2\log x)^{\frac{-1}{2}}\cos(2\pi y) \\
\frac{\partial h_1}{\partial y} &= (-2\pi)(-2\log x)^{\frac{-1}{2}}\sin(2\pi y) \\
\frac{\partial h_2}{\partial x} &= \big(\frac{1}{2}\big)\big(\frac{-2}{x}\big)(-2\log x)^{\frac{-1}{2}}\sin(2\pi y) \\
\frac{\partial h_2}{\partial y} &= (2\pi)(-2\log x)^{\frac{-1}{2}}\cos(2\pi y)
\end{aligned}$$

these are all continuous functions of x and y in $R_{XY} = \{(x, y) \mid 0 < x, y < 1\}$ thus by the inverse mapping theorem we have a 1-to-1 function.

Step 2: find determinant of Jacoby

$$\begin{aligned}
\frac{\partial(h_1, h_2)}{\partial(x, y)} &= \frac{\partial h_1}{\partial x}\frac{\partial h_2}{\partial y} - \frac{\partial h_1}{\partial y}\frac{\partial h_2}{\partial x} \\
&\vdots \\
&= \frac{-2\pi}{x}
\end{aligned}$$

which is $\neq 0$ in R_{XY}.

Note that $u^2 + v^2 = -2\log x = w_1(u, v)$ thus $x = e^{-\frac{1}{2}(u^2+v^2)}$ so we have

$$\begin{aligned}
\left|\frac{\partial(w_1, w_2)}{\partial(u, v)}\right| &= \left|\big(\frac{\partial(h_1, h_2)}{\partial(x, y)}\big)^{-1}\right| \\
&= \frac{x}{2\pi} \\
&= \frac{e^{-\frac{1}{2}(u^2+v^2)}}{2\pi}
\end{aligned}$$

(note in this case an explicit $y = w_2(u, v)$ is difficult to derive and is also not 1-to-1 since cos and sin are not 1-to-1).

Step 3: find $g(u, v)$ We have

$$\begin{aligned}
g(u, v) &= \frac{e^{-\frac{1}{2}(u^2+v^2)}}{2\pi} \\
&= \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{2}}
\end{aligned}$$

which is the product of two $N(0, 1)$ (by the factorization theorem we have independence between U and V).

Step 4: find R_{UV} Note previously we found R_{UV} from R_{XY} by considering $x = w_1(u, v)$ and $y = w_2(u, v)$, however we do not have an explicit w_2.

We observe what happens to our functions h_1 and h_2:

1. When $0 < x < 1$, we have $(-2\log x)^{\frac{1}{2}} > 0$.
2. When $0 < y < 1$ we have $-1 \leq \cos(2\pi y), \sin(2\pi y) \leq 1$

Therefore $R_{UV} = \{(u,v) \mid u,v \in \mathbb{R}\}$ where U,V are independent by the factorization theorem, $U,V \sim N(0,1)$.

Remark 16.1. Since U,V are functions of uniform r.v.s X,Y, this tells us how to generate independent normal r.v.s from independent uniform r.v.s.

17 October 22, 2018

17.1 Univariate transformation with mgf technique

Let $M_X(t)$ be the mgf of X for $|t| < h$, some $h > 0$.
Let $Y = aX + b$, $a \neq 0, b \in \mathbb{R}$. Find $M_Y(t)$.
Note that

$$M_Y(t) = E[e^{tY}] = E[e^{t(aX+b)}] = e^{tb}E[e^{(ta)X}]$$

Let $t^* = ta$ then $M_Y(t) = e^{tb}E[e^{t^*X}]$ where $E[e^{t^*X}] = M_X(t^*)$, thus $M_Y(t) = e^{tb}M_X(ta)$. We need to write bounds for t in $M_Y(t)$: that is $|ta| < h$ iff $|t| < \frac{h}{|a|}$.
Some special results

1. If $X \sim GAM(\alpha, \beta)$ and α is a positive integer, let $Y = \frac{2X}{\beta}$. Then $Y \sim \chi^2(2\alpha)$ (chi-squared).

   Proof. We have $Y = \frac{2X}{\beta}$ i.e. $a = \frac{2}{\beta}$, $b = 0$ (notation from our univariate linear transformation example).
   Thus

$$\begin{aligned} M_Y(t) &= e^{t\cdot 0}M_X(t\cdot\frac{2}{\beta}) \\ &= M_X(\frac{2t}{\beta}) \\ &= M_X(t^*) && t^* = \frac{2t}{\beta} \end{aligned}$$

   Recall that if $X \sim \Gamma(\alpha, \beta)$ then $M_X(t) = \frac{1}{(1-\beta t)^\alpha}$, $t < \frac{1}{\beta}$.
   Thus

$$\begin{aligned} M_Y(t) = M_X(t^*) &= \frac{1}{(1-\beta t^*)^\alpha} && t^* < \frac{1}{\beta} \\ &= \frac{1}{(1-\beta\cdot\frac{2t}{\beta})^\alpha} && \frac{2t}{\beta} < \frac{1}{\beta} \\ &= \frac{1}{(1-2t)^{\frac{2\alpha}{2}}} && t < \frac{1}{2} \end{aligned}$$

   Note that if $X \sim \chi^2(n)$, then $M_X(t) = \frac{1}{(1-2t)^{\frac{n}{2}}}$, $t < \frac{1}{2}$.

   Thus $Y \sim \chi^2(2\alpha)$ due to the uniqueness theorem of mgfs.

   If $X_i \sim GAM(\alpha_i, \beta)$, $i = 1,\ldots,n$ independent, then $\sum_{i=1}^n X_i \sim GAM(\sum_{i=1}^n \alpha_i, \beta)$.

Proof.

$$\begin{aligned}
M_Y(t) = E[e^{tY}] &= E[e^{t\sum_{i=1}^n X_i}] \\
&= \prod_{i=1}^n E[e^{tX_i}] && \text{independence} \\
&= \frac{1}{\prod_{i=1}^n (1-\beta t)^{\alpha_i}} && t < \frac{1}{\beta} \\
&= \frac{1}{(1-\beta t)^{\sum_{i=1}^n \alpha_i}}
\end{aligned}$$

which is the mgf of $GAM(\sum_{i=1}^n \alpha_i, \beta)$. □

□

2. If $X_i \sim EXP(\beta)$, then $\sum_{i=1}^n X_i \sim GAM(n, \beta)$.

   Proof. Exercise (hint: $EXP(\beta) \sim GAM(1, \beta)$). □

3. If $X_i \sim \chi^2(k_i)$, $i = 1, \dots, n$ independent, then $\sum_{i=1}^n X_i \sim \chi^2(\sum_{i=1}^n k_i)$.

   Proof. Similar to sum of Gamma proof. □

4. If $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, then $\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2 = \chi^2(n)$.

   Proof. Hint: note $\frac{x_i-\mu}{\sigma} \sim N(0,1)$ and $N(0,1)^2 \sim \chi^2(1)$. □

5. If $X_i \overset{iid}{\sim} POI(\mu_i)$, then $\sum_{i=1}^n X_i \sim POI(\sum_{i=1}^n \mu_i)$.

6. If $X_i \overset{iid}{\sim} BIN(n_i, p)$, then $\sum_{i=1}^n X_i \sim BIN(\sum_{i=1}^n n_i, p)$.

17.2 Sum and mean of Gaussian random variables

Theorem 17.1. If $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$ independent, then

$$\sum_{i=1}^n a_i X_i \sim N(\sum_{i=1}^n a_i X_i, \sum_{i=1}^n a_i^2 \sigma_i^2)$$

Proof. Exercise. □

Corollary 17.1. If $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

Proof. Let $a_i = \frac{1}{n}$. Then we have $\sum_{i=1}^n \frac{1}{n} X_i = \bar{X}$. From our theorem above we have

$$\begin{aligned}
\bar{X} &\sim N(\sum_{i=1}^n \frac{1}{n}\mu, \sum_{i=1}^n \frac{1}{n^2}\sigma^2) \\
&\sim N(\mu, \frac{\sigma^2}{n})
\end{aligned}$$

□

18 October 24, 2018

18.1 Independence of mean and sample variance

Theorem 18.1. If $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ is independent of $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ where $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ is the sample variance.

Proof. Outline of steps to complete proof:

1. $\bar{X}$ and $(n-1)S^2$ are independent (by independent theorem of mgfs)

2. $$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} = \frac{n(\bar{X} - \mu)^2}{\sigma^2} + \frac{(n-1)S^2}{\sigma^2}$$

   thus we have $\chi^2(n) = \chi^2(1) + \chi^2(n-1)$.

3. $$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

   from special result 14.

1. Note that $(n-1)S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = (X_1 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2$ i.e. $(n-1)S^2$ is a function of $\{(X_1 - \bar{X}), \ldots, (X_n - \bar{X})\}$.

   To show $\bar{X}$ is independent of $(n-1)S^2$, it suffices to show $\bar{X}$ is independent of $\{(X_1 - \bar{X}), \ldots, (X_n - \bar{X})\}$.

   Let $U_i = X_i - \bar{X}$. Find the joint mgf of $(U_1, U_2, \ldots, U_n, \bar{X})$ ($n+1$ entries), and then the marginal mgfs of $(U_1, \ldots, U_n)$ and $\bar{X}$ respectively.

   We have

   $$\begin{aligned} M(s_1, \ldots, s_n, s_0) &= E[e^{(s_1, \ldots, s_n, s_0)(U_1, \ldots, U_n, \bar{X})^T}] \\ &= E[e^{s_0 \bar{X} + \sum_{i=1}^n s_i U_i}] \end{aligned}$$

   Notice that $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$ with a common mgf $e^{t\mu + \frac{1}{2}t^2\sigma^2}$, $t \in \mathbb{R}$.

   Also $\bar{X} = \frac{1}{n}\sum_{n=1}^n X_i$ and $U_i = X_i - \bar{X}$. Ideally we want to decompose our joint mgf into a product of the marginal mgfs: re-arranging we have

$$\begin{aligned}
s_0\bar{X} + \sum_{i=1}^n s_i U_i &= s_0 + \sum_{i=1}^n s_i(X_i - \bar{X}) \\
&= (s_0 - \sum_{i=1}^n s_i)\bar{X} + \sum_{i=1}^n s_i X_i \\
&= (s_0 - \sum_{i=1}^n s_i)\frac{1}{n}\sum_{i=1}^n X_i + \sum_{i=1}^n s_i X_i \\
&= (s_0 - \sum_{i=1}^n s_i)\frac{1}{n}\sum_{i=1}^n X_i + \sum_{i=1}^n s_i X_i \\
&= \sum_{i=1}^n (\frac{s_0}{n} - \frac{\sum_{i=1}^n s_i}{n})X_i + \sum_{i=1}^n s_i X_i \\
&= \sum_{i=1}^n (\frac{s_0}{n} - \bar{s} + s_i)X_i
\end{aligned}$$

Let $t_i = \frac{s_0}{n} - \bar{s} + s_i$, $i = 1, \ldots, n$ therefore we have

$$\begin{aligned}
M(s_1, \ldots, s_n, s_0) &= E[e^{\sum_{i=1}^n t_i X_i}] \\
&= \prod_{i=1}^n E[e^{t_i X_i}] && \text{independence} \\
&= \prod_{i=1}^n e^{t_i\mu + \frac{1}{2}t_i^2\sigma^2} \\
&= e^{\mu\sum_{i=1}^n t_i + \frac{1}{2}\sigma^2\sum_{i=1}^n t_i^2}
\end{aligned}$$

Note that we have

$$\begin{aligned}
\sum_{i=1}^n t_i &= \sum_{i=1}^n \frac{s_0}{n} - \bar{s} + s_i \\
&= s_0 - n\bar{s} + \sum_{i=1}^n s_i \\
&= s_0 && \sum_{i=1}^n s_i = n\bar{s}
\end{aligned}$$

also

$$\begin{aligned}
\sum_{i=1}^n t_i^2 &= \sum_{i=1}^n \left(\frac{s_0}{n}(s_i - \bar{s})\right)^2 \\
&= \sum_{i=1}^n \big(\frac{s_0^2}{n^2} + (s_i - \bar{s})^2 - 2\frac{s_0}{n}(s_i - \bar{s})\big) \\
&= \frac{s_0^2}{n} + \sum_{i=1}^n (s_i - \bar{s})^2 - 2\frac{s_0}{n}\sum_{i=1}^n (s_i - \bar{s}) \\
&= \frac{s_0^2}{n} + \sum_{i=1}^n (s_i - \bar{s})^2
\end{aligned}$$

Therefore we have

$$\begin{aligned}
M(s_1, \ldots, s_n, s_0) &= e^{\mu \sum_{i=1}^n t_i + \frac{1}{2}\sigma^2 \sum_{i=1}^n t_i^2} \\
&= e^{\mu s_0 + \frac{1}{2}\sigma^2 \left(\frac{s_0^2}{n} + \sum_{i=1}^n (s_i - \bar{s})^2\right)} \\
&= e^{\mu s_0 + \frac{1}{2}\sigma^2 \frac{s_0^2}{n}} e^{\frac{1}{2}\sigma^2 \sum_{i=1}^n (s_i - \bar{s})^2}
\end{aligned}$$

for $t_i \in \mathbb{R}$ and $t_i = \frac{s_0}{n} + s_i - \bar{s}$, therefore $s_i \in \mathbb{R}$ and $s_0 \in \mathbb{R}$.

Note that $M_{\bar{X}}(s_0) = M(s_1, \ldots, s_n = 0, s_0) = e^{\mu s_0 + \frac{1}{2}s_0^2 \frac{\sigma^2}{n}}$ which is identical to the mgf of $N(\mu, \frac{\sigma^2}{n})$, therefore $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ (confirming our previous corollary).

Also $M_{U_1,\ldots,U_n}(s_1, \ldots, s_n, 0) = e^{\frac{1}{2}\sigma^2 \sum_{i=1}^n (s_i - \bar{s})^2}$.

Thus we have

$$M(s_1, \ldots, s_n, s_0) = M_{\bar{X}}(s_0) \cdot M_{U_1,\ldots,U_n}(s_1, \ldots, s_n)$$

so $\bar{X}$ and $(U_1, \ldots, U_n)$ are independent due to the mgf independence theorem.

2. We want to show

$$\begin{aligned}
\sum_{i=1}^n (X_i - \mu)^2 &= (n-1)S^2 + n(\bar{X} - \mu)^2 \\
&= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2
\end{aligned}$$

Note that

$$
\begin{aligned}
LHS &= \sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n} \big((X_i - \bar{X})^2 + (\bar{X} - \mu)^2 + 2(\bar{X} - \mu)(X_i - \bar{X}) \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 + 2(\bar{X} - \mu)\sum_{i=1}^{n}(X_i - \bar{X}) \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2
\end{aligned}
$$

Therefore we have

$$\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2}$$

3. We want to show

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

First

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$$

(from special result above).

Secondly,

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

so

$$
\begin{aligned}
&\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1) \\
\Rightarrow &\left(\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}\right)^2 \sim \chi^2(1)
\end{aligned}
$$

where

$$\left(\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}\right)^2 = \frac{n(\bar{X} - \mu)^2}{\sigma^2} \sim \chi^2(1)$$

Therefore from step two we have

$$\sum_{i=1}^{n}\frac{(X_i - \mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2}$$

from step 1 we know the right two terms are independent. The LHS is $X^2(n)$ and the the right term on the RHS is $X^2(n-1)$ thus

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

□

19 October 26, 2018

19.1 T distribution

Theorem 19.1. Suppose $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$. Then

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$$

Proof. We define the t distribution as: if $Z \sim N(0,1)$ independent of $X \sim \chi^2(n)$ then

$$\frac{Z}{\sqrt{\frac{X}{n}}} \sim t(n)$$

To prove the statement we need to show that

$$T = \frac{N(0,1)}{\sqrt{\frac{\chi^2(n-1)}{n-1}}}$$

Recall that $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ i.e. $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$. Furthermore $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ and $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ and $\frac{(n-1)S^2}{\sigma^2}$ are independent therefore

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{N(0,1)}{\sqrt{\frac{\chi^2(n-1)}{n-1}}}$$

as desired. □

19.2 F distribution

Theorem 19.2. Suppose $X_i \overset{iid}{\sim} N(\mu, \sigma_1^2)$ and $Y_i \overset{iid}{\sim} N(\mu, \sigma_2^2)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.
Let $S_1^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ and $S_2^2 = \frac{1}{n-1}\sum_{i=1}^m (Y_i - \bar{Y})^2$. Show that

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim f(n-1, m-1)$$

Proof. Recall that $\frac{\chi^2(n)/n}{\chi^2(m)/m} \sim f(n,m)$ as long as numerator and denominator are independent.
We know that $\frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2(n-1)$ and $\frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2(m-1)$, and they are independent since they are from two random samples (i.e. $(X_1, \ldots, X_n)$ are independent of $(Y_1, \ldots, Y_n)$).
Therefore

$$\frac{\frac{(n-1)S_1^2}{\sigma_1^2}/(n-1)}{\frac{(m-1)S_2^2}{\sigma_2^2}/(m-1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim f(n-1, m-1)$$

□

19.3 Limiting distributions and asymptotics

Question 19.1. What is the limiting behavior of a set of random variables $X_1, \ldots, X_n$ as $n \to \infty$?

Later on we will look at the **Weak Law of Large Numbers (WLLN)** for convergence in probability and the **Central Limit Theorem (CLT)** for convergence in distributions.
Let $X_1, \ldots, X_n$ be a sequence of random variables with cdf $F_1(x), \ldots, F_n(x)$, respectively.
Let X be another r.v. with cdf $F(x)$.

Convergence in probability

Theorem 19.3. Convergence in probability $X_n \xrightarrow{P} X$ if for any $\epsilon > 0$ (however small)

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$$

or

$$\lim_{n\to\infty} P(|X_n - X| \leq \epsilon) = 1$$

(notice $P(|X_n - X| > \epsilon) = 1 - P(|X_n - X| \leq \epsilon)$, therefore the above are equivalent).

That is: For large n, X_n is close to X with probability approaching to 1.

Convergence in distribution

Theorem 19.4. Convergence in distribution $X_n \xrightarrow{D} X$ if $\lim_{n\to\infty} F_n(x) = F(x)$ for all x where $F(\cdot)$ are continuous.

Remark 19.1. If we are interested in calculating $f_n(x)$ for large n, we can calculate $f(x)$ (pf of X) instead if x is a continuity point of $f(\cdot)$.

Remark 19.2. When verifying $X_n \xrightarrow{D} X$, we only focus on continuous points of $F(\cdot)$: it is okay if $\lim_{n\to\infty} F_n(x) \neq F(x)$ where $F(\cdot)$ is not continuous.

Example 19.1. Suppose $X_n \sim N(0, \frac{1}{n})$, $n = 1, 2, \ldots$. What is the limiting distribution of X_n?
Notice that $N(0, \frac{1}{n})$ is symmetric around mean 0 as $n \to \infty$.

Question 19.2. Is there convergence in the distribution, in probability, or both?

First: let's look at the convergence in distribution. Note

$$\begin{aligned}
F_n(x) &= P(X_n \leq x) & \\
&= P(\sqrt{n}X_n \leq \sqrt{n}x) & \qquad \frac{X_n}{\sqrt{1/n}} \sim N(0,1) \\
&= \Phi(\sqrt{n}x) & \qquad \Phi \text{ cdf of } N(0,1)
\end{aligned}$$

Thus

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} \Phi(\sqrt{n}x) = \begin{cases} \Phi(\infty) = 1 & \text{if } x > 0 \\ \Phi(0) = \frac{1}{2} & \text{if } x = 0 \\ \Phi(-\infty) = 0 & \text{if } x < 0 \end{cases}$$

This is not possibly a cdf as $\lim_{n\to\infty} F_n(x)$ is not right continuous at $x = 0$.
Now let $X = 0$, a degenerate r.v.
$P(X = 0) = 1$ so the cdf of X is $F(x) = 0$ if $x < 0$ and $F(x) = 1$ if $x \geq 0$. So $\lim_{n\to\infty} F_n(x) = F(x)$ at all x where $F(\cdot)$ is continuous ($F(\cdot)$ is only not continuous at $x = 0$).

20 October 29, 2018

20.1 Relationship between convergence in probability and distribution

Remark 20.1. For verifying $X_n \xrightarrow{P} X$ is often done by Markov's inequality where we attempt to show

$$0 \leq \lim_{n\to\infty} P(|X_n - X| > \epsilon) \leq 0$$

from the Squeeze theorem $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$.

We remark that

1. $X_n \xrightarrow{P} X$ implies $X_n \xrightarrow{D} X$
2. $X_n \xrightarrow{D} X = c$ implies $X_n \xrightarrow{P} X = c$

Remark 20.2. In general $X_n \xrightarrow{D} X$ does not imply $X_n \xrightarrow{P} X$, unless $X = c$ (a real constant), then $X_n \xrightarrow{D} X = c$ is equivalent to $X_n \xrightarrow{P} X = c$.

Remark 20.3. Convergence in probability is a stronger convergence (and convergence in distribution is a weaker convergence).

We will look at the case where convergence in distribution does not imply convergence in probability.

Example 20.1. Suppose $X \sim N(0,1)$ and $X_n = -X$, $n = 1, 2, \ldots$ i.e. $X_1, X_2, \ldots$ has the same distribution where $F_1(x) = F_2(x) = \ldots$.
Notice that $X \sim N(0,1)$, is a symmetric (around 0) r.v. so $-X \sim N(0,1)$.
Therefore $F_1(x) = F_2(x) = \ldots = \Phi(x)$ (Φ is cdf of $N(0,1)$) for all $x \in \mathbb{R}$ since

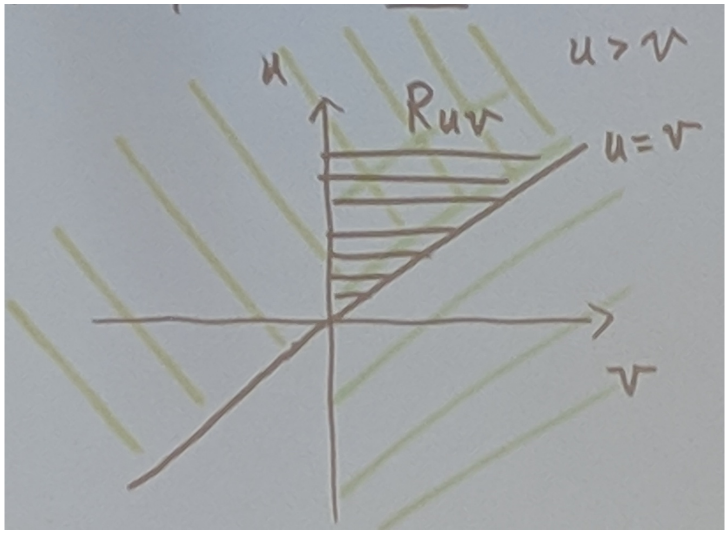
$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} \Phi(x) = \Phi(x) \qquad \forall x \in \mathbb{R}$$

i.e. $X_n \xrightarrow{D} X$. Notice $\Phi(\cdot)$ is cont. at any $x \in \mathbb{R}$.
But notice $X_n \overset{P}{\not\to} X$ since

$$\begin{aligned}
P(|X_n - X| > \epsilon) &= P(|-X - X| > \epsilon) \\
&= P(2|X| > \epsilon) \\
&= P(|X| > \frac{\epsilon}{2}) \\
&= P(X > \frac{\epsilon}{2}) + P(X < \frac{-\epsilon}{2}) \\
&= 2\Phi(-\frac{\epsilon}{2})
\end{aligned}$$

20.2 More bivariate transformation examples

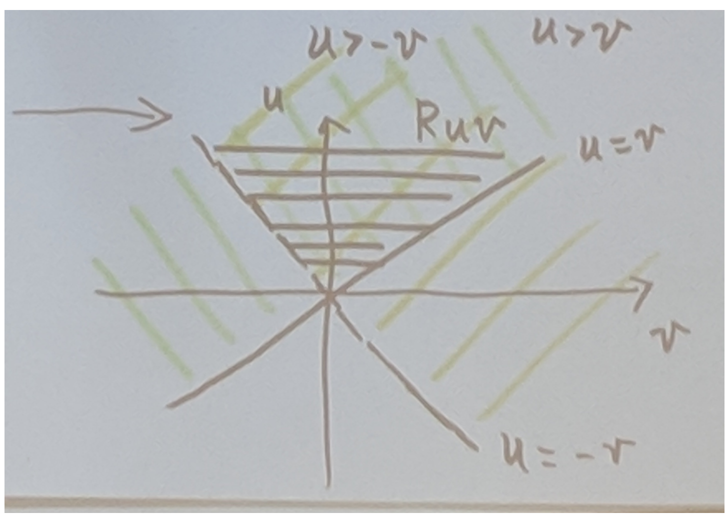
Exercise 20.1. Suppose $g(u, v) = e^{-u}$ and $R_{UV} = \{(u, v) \mid v > 0, u - v > 0\}$.



We can rewrite the support set to help us find the marginal pdfs i.e. $R_{UV} = \{(u,v) \mid u > 0, 0 < v < u\}$. Thus to find each of the marginals, we simply integrate over the other variable

$$g_1(u) = \int_0^v g(u,v)\,\mathrm{d}v \qquad u > 0$$
$$g_2(v) = \int_v^\infty g(u,v)\,\mathrm{d}u \qquad v > 0$$

Exercise 20.2. Suppose $g(u,v) = e^{-u/2}$ and $R_{UV} = \{(u,v) \mid u+v > 0, u-v > 0\}$, so $u > -v$ and $u > v$.



We can rewrite the support set to help us find the marginal pdfs i.e.

$$\begin{aligned} R_{UV} &= \{(u,v) \mid u > 0, -u < v < u\} \\ &= \{(u,v) \mid v < 0, u > -v\} \cup \{(u,v) \mid v \geq 0, u > v\} \end{aligned}$$

Thus to find each of the marginals, we simply integrate over the other variable

$$
\begin{aligned}
g_1(u) &= \int_{-u}^{u} g(u,v)\,\mathrm{d}v & u > 0 \\
g_2(v) &= \int_{0}^{\infty} g(u,v)\,\mathrm{d}u & v < 0 \\
g_2(v) &= \int_{0}^{\infty} g(u,v)\,\mathrm{d}u & v > 0
\end{aligned}
$$

Exercise 20.3. Show that

$$\frac{X/n}{Y/m} \sim F(n,m)$$

if $X \sim \chi^2(n)$ independent of $Y \sim \chi^2(m)$.

Proof. Let $U = \frac{X/n}{Y/m}$, $V = Y$.
We have

$$g(u,v) = K \cdot u^{\frac{n}{2}-1} e^{-v(\frac{1}{2}+\frac{nu}{2m})} v^{\frac{n+m}{2}-1}$$

where $R_{UV} = \{(u,v) \mid u > 0, v > 0\}$ and $g_1(u) = \int_0^\infty g(u,v)\,\mathrm{d}v$. We end up with

$$
\begin{aligned}
g_1(u) &= K \cdot u^{\frac{n}{2}-1} \int_0^\infty e^{-v(\frac{1}{2}+\frac{nu}{2m})} v^{\frac{n+m}{2}-1}\,\mathrm{d}v \\
&= K \cdot u^{\frac{n}{2}-1} \beta^\alpha \Gamma(\alpha) \\
&= K \cdot u^{\frac{n}{2}-1} (\frac{1}{2} + \frac{nu}{2m})^{-\frac{n+m}{2}} \Gamma(\frac{n+m}{2})
\end{aligned}
$$

which is the pdf of $F(n,m)$. □

21 October 31, 2018

21.1 Example of convergence in distribution

Example 21.1. Suppose $X_i \overset{iid}{\sim} EXP(1)$, $i = 1, \ldots, n$ for a given n.
Let $Y_n = \max(X_1, \ldots, X_n) - \log n$ for $n = 1, 2, \ldots$.
Show $Y_n \xrightarrow{D} Y$ and identify Y.

Solution. Let

$$
\begin{aligned}
F_n(y) &= P(Y_n < y) \\
&= P(\max(X_1, \ldots, X_n) - \log n \leq y) \\
&= P(\max(X_1, \ldots, X_n) \leq y + \log n) \\
&= \prod_{i=1}^{n} P(X_i \leq y + \log n)
\end{aligned}
$$

Recall if $X \sim EXP(1)$ then the cdf of X is

$$P(X \leq x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - e^{-x} & \text{if } x > 0 \end{cases}$$

Thus

$$\begin{aligned} F_n(y) &= \big(P(X_1 \le y + \log n)\big)^n \\ &= \begin{cases} 0 & \text{if } y + \log n \le 0 \\ \big(1 - e^{-(y+\log n)}\big)^n & \text{if } y + \log n > 0 \end{cases} \end{aligned}$$

Note that

$$\lim_{n\to\infty} F_n(y) = \begin{cases} 0 & \text{if } y \le -\log n \overset{n\to\infty}{\Rightarrow} y = -\infty \\ \lim_{n\to\infty} \big(1 - e^{-(y+\log n)}\big)^n & \text{if } y > -\infty \end{cases}$$

Notice that

$$\lim_{n\to\infty} \big(1 - e^{-y} e^{-\log n)}\big)^n = \lim_{n\to\infty} \big(1 - \frac{e^{-y}}{n}\big)^n$$

Recall that

$$\lim_{n\to\infty} = (1 + \frac{b}{n})^n = e^b \qquad b \in \mathbb{R}$$

thus we have

$$\lim_{n\to\infty} \big(1 - \frac{e^{-y}}{n}\big)^n = e^{-e^{-y}}$$

So in summary

$$\lim_{n\to\infty} F_n(y) = \begin{cases} 0 & \text{if } y \le -\log n \overset{n\to\infty}{\Rightarrow} y = -\infty \\ e^{-e^{-y}} & \text{if } y > -\infty \end{cases}$$

note that $\lim_{y\to-\infty} e^{-e^{-y}} = 0$ so we simply have $\lim_{n\to\infty} F_n(y) = e^{-e^{-y}}$ for all $y \in \mathbb{R}$, that is $Y_n \overset{D}{\to} Y$ with a cdf $F(y) = e^{-e^{-y}}$, $y \in \mathbb{R}$.
In fact $(-Y) \sim EV(1, 0)$: the **extreme value distribution**.

21.2 Example of convergence in probability

Example 21.2. Suppose $X_i \overset{iid}{\sim} UNIF(0, \theta)$, $\theta > 0$. Let $Y_n = \max(X_1, \dots, X_n)$. Show $Y_n \overset{P}{\to} Y$ and identify Y.

Solution. Intuition: as $n \to \infty$, we have more and more uniform sample points on the interval $(0, \theta)$. Y_n is the maximum value of all these sample points so we expect $\lim_{n\to\infty} P(Y_n = \theta) = 1$ i.e. $Y_n \overset{P}{\to} Y = \theta$.

Method 1 Show that $Y_n \overset{D}{\to} Y = \theta$, which implies $Y_n \overset{P}{\to} Y = \theta$.

Let $F_n(y) = P(max(X_1, \dots, X_n) \le y) = P(X_i \le y)^n$.

Note that for $X \sim UNIF(0, \theta)$

$$P(X \le x) = \begin{cases} 0 & \text{if } x \le 0 \\ \frac{x}{\theta} & \text{if } 0 < x < \theta \\ 1 & \text{if } x \ge \theta \end{cases}$$

Thus we have

$$F_n(y) = \begin{cases} 0 & \text{if } y \le 0 \\ \big(\frac{y}{\theta}\big)^n & \text{if } 0 < y < \theta \\ 1 & \text{if } y \ge \theta \end{cases}$$

Thus we have

$$\lim_{n\to\infty} F_n(y) = \begin{cases} 0 & \text{if } y < \theta \\ 1 & \text{if } y \geq \theta \end{cases}$$

since $\frac{y}{\theta} < 1$. Note that $Y = \theta$ has the exact same cdf so $\lim_{n\to\infty} F_n(y) = F(y)$ for all $y \in \mathbb{R}$ (includes all continuity points of $F(\cdot)$). Therefore $Y_n \xrightarrow{D} Y$ so $Y_n \xrightarrow{P} Y$.

Method 2 By definition of convergence in probability (left as an exercise).

Remark 21.1. Let Y_n be an estimator of θ: Y_n should be a good estimator of θ as $Y_n \xrightarrow{P} Y = \theta$ i.e. Y_n is close to 0 with probability approaching 1 for large n.

21.3 Sequence of maximum of uniform r.v.s

Theorem 21.1. Suppose X_n has a cdf $F_n(x)$, $n = 1, 2, \ldots$. If

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & \text{if } x < b \\ 1 & \text{if } x \geq b \end{cases} \qquad \forall b \in \mathbb{R}$$

then $X_n \xrightarrow{P} X = b$.

Proof. Suppose

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{if } x < b \\ 1 & \text{if } x = b \\ 1 & \text{if } x > b \end{cases}$$

Therefore $\lim_{n\to\infty} F_n(x) = F(x)$ at all continuity points of $F(\cdot)$ (notice that $x = b$ is the only discontinuous point of $F(\cdot)$).
So $X_n \xrightarrow{D} X = b$ then $X_n \xrightarrow{P} X = b$. □

21.4 Sequence of minimum of shifted exponential r.v.s

Theorem 21.2. Let $X_i \overset{iid}{\sim} EXP(1, \theta)$ and let $Y_n = \min(X_1, \ldots, X_n)$. Then $Y_n \xrightarrow{P} Y = \theta$.

Proof. Recall that $X \sim EXP(1, \theta)$ has pdf $e^{x-\theta}$ for $x \geq \theta$.
So we expect $Y_n \xrightarrow{P} Y = \theta$ i.e. it approaches the lower bound of all the X_i's.

Method 1 Show $Y_n \xrightarrow{P} Y = \theta$ using theorem 5.2.4 (left as an exercise).

Method 2 Using the definition of convergence in probability we show $P(|Y_n - Y| > \epsilon)$ for any $\epsilon > 0$.

$$\begin{aligned} P(|Y_n - \theta| > \epsilon) &= P(|min(X_1, \ldots, X_n) - \theta| > \epsilon) \\ &= P(min(X_1, \ldots, X_n) - \theta > \epsilon) && X_i \geq \theta \\ &= P(min(X_1, \ldots, X_n) > \epsilon + \theta) \\ &= \prod_{i=1}^{\infty} P(X_i > \epsilon + \theta) \\ &= \big(P(X_1 > \epsilon + \theta)\big)^n \end{aligned}$$

Recall for $X \sim EXP(1,\theta)$ the cdf is

$$P(X \leq x) = \begin{cases} 0 & \text{if } x < \theta \\ 1 - e^{-(x-\theta)} & \text{if } x \geq \theta \end{cases}$$

So

$$P(X_1 > \theta + \epsilon) = 1 - P(X_1 \leq \theta + \epsilon) = 1 - (1 - e^{-(\theta+\epsilon-\theta)}) = e^{-\epsilon}$$

In summary $P(|Y_n - Y| > \epsilon) = (e^{-\epsilon})^n = e^{-n\epsilon}$, so $\lim_{n\to\infty} P(|Y_n - Y| > \epsilon) = \lim_{n\to\infty} e^{-n\epsilon} = 0$.

Thus $Y_n \xrightarrow{P} Y = \theta$.

□

Remark 21.2. We can use a sequence of $X_i \sim EXP(1,\theta)$ and $Y_n = \min(X_1, \ldots, X_n)$ to estimate θ.

22 November 2, 2018

22.1 Mgf limit theorem

Theorem 22.1 (Mgf limit theorem)**.** Let $M_n(t)$ and $M(t)$ be the mgfs of X_n and X, respectively where $n = 1, 2, \ldots$. Then $X_N \xrightarrow{D} X$ if $\lim_{n\to\infty} M_n(t) = M(t)$ for all $t \in (-h, h)$, some $h > 0$.

Remark 22.1. Suppose we manage to verify that $\lim_{n\to\infty} M_n(t) = M(t)$ for $t < \frac{1}{2}$.
By the theorem we need to find a neighbourhood with radius h around 0, which we can do for $h = \frac{1}{4}$ i.e. the limit holds for $t \in (-1/4, 1/4)$ so $X_n \xrightarrow{D} X$ holds.
Similarly if the limit holds for $t \in \mathbb{R}$, then we can definitely find a neighbourhood around 0 (in fact neighbourhood of radius h for any $h > 0$).

22.2 Poisson approximation to Binomial (example of mgf limit theorem)

Example 22.1. Suppose $X_n \sim BIN(n, p)$ where $p = \frac{\mu}{n}$ such that $0 < \frac{\mu}{n} < 1$ for $n = 1, 2, \ldots$.
Show that $X_n \xrightarrow{D} POI(\mu)$.

Proof. Let $M_n(t)$ be the mgf of X_n, $n = 1, 2, \ldots$ and let $M(t)$ be the mgf of $X \sim POI(\mu)$ i.e. $M(t) = e^{\mu(e^t-1)}$.
To show $X_n \xrightarrow{D} X \sim POI(\mu)$, it suffices to show $\lim_{n\to\infty} M_n(t) = M(t) = e^{\mu(e^t-1)}$ for all $t \in (-h, h)$ for some $h > 0$.
$M_n(t) = E[e^{tX_n}]$ where $X_n \sim BIN(n, \frac{\mu}{n})$.
Recall for $Y \sim BIN(n, p)$ we have $M_Y(t) = (e^t p + 1 - p)^n$, $t \in \mathbb{R}$.
Thus $M_n(t) = (e^t \frac{\mu}{n} + 1 - \frac{\mu}{n})^n$ for $t \in \mathbb{R}$. Taking the limit we get

$$\begin{aligned} \lim_{n\to\infty} M_n(t) &= \lim_{n\to\infty} (1 + \frac{\mu(e^t-1)}{n})^n \\ &= e^{\mu(e^t-1)} & (1 + \frac{b}{n})^n = e^b \end{aligned}$$

for any $t \in \mathbb{R}$, thus by the mgf limit theorem we have $X_n \xrightarrow{D} X \sim POI(\mu)$. □

Remark 22.2. Note that $P(POI(\mu) = r) \approx P(BIN(n, \frac{\mu}{n}) = r)$ for $r = 0, 1, 2, \ldots$ for a large n.
Since $P(POI(\mu) = r)$ is easier to compute so we can approximate a binomial from a Poisson distribution.
Notice that from $X_n \xrightarrow{D} X$ we only have $\lim_{n\to\infty} F_n(x) = F(x)$ for all continuous points of $F(\cdot)$ i.e. we only have a convergence in the cdf. How do we show that the pmf also converges?

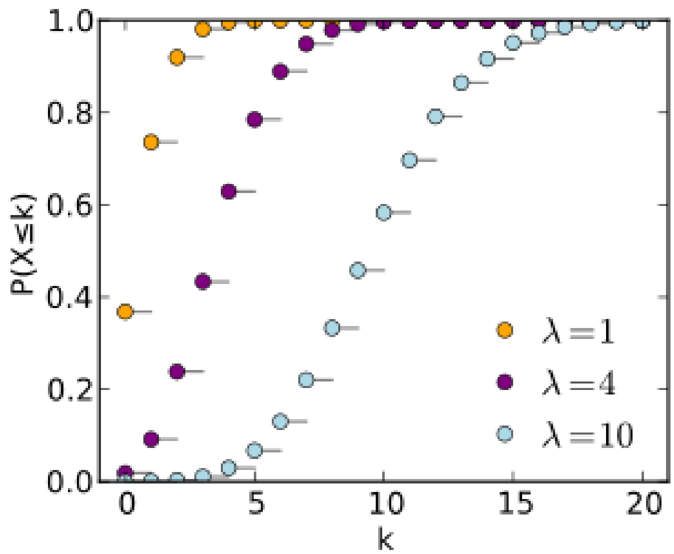


Figure 22.1: CDF of $X \sim POI(\mu)$.

Proof. Note that $F(x)$ is continuous for any $x \in \mathbb{R}$ except at $x = 0, 1, 2, \ldots$.
Therefore $\lim_{n\to\infty} F_n(x) = F(x)$ for all $x \in \mathbb{R}$ except at $x = 0, 1, 2, \ldots$.
We transform the cdf values to our desired pmf:

cdf of $X_n \sim BIN(n, \mu/n)$ Note that

$$\begin{aligned} F_n(0.1) - F_n(-0.1) &= P(X_n \leq 0.1) - P(X_n \leq -0.1) \\ &= P(X_n = 0) - 0 \\ &= P(X_n = 0) \end{aligned}$$

cdf of $X \sim POI(\mu)$ Note that

$$\begin{aligned} F(0.1) - F(-0.1) &= P(X \leq 0.1) - P(X \leq -0.1) \\ &= P(X = 0) - 0 \\ &= P(X = 0) \end{aligned}$$

Due to convergence in distributions we have $\lim_{n\to\infty} F_n(0.1) = F_n(0.1)$ (and similarly with $F_n(-0.1)$ and $F(-0.1)$), so $\lim_{n\to\infty} P(X_n = 0) = P(X = 0)$ i.e. $P(X = 0) \approx P(X_n = 0)$ for a large n. ☐

22.3 Central limit theorem

Theorem 22.2 (Central limit theorem)**.** Suppose X_n are iid with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Then $Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} Z \sim N(0, 1)$ where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Remark 22.3. In our proof for the mean of Gaussian r.v.s we showed this holds for $X_i \sim N(\mu, \sigma^2)$.

Remark 22.4. We will show this via the mgf limit theorem (which requires more assumptions than necessary for the CLT i.e. we require X_n's to have mgf $M_n(t)$ for all $t \in (-h, h)$ for some $h > 0$).
Note that $E(X_n^r) < \infty$ for $r = 1, 2, \ldots$. If we prove the CLT using the characteristic function then we do not require the additional assumptions for the mgf limit theorem).

Proof. To show $Z_n = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \xrightarrow{D} Z \sim N(0,1)$ via the the mgf limit theorem, it suffices to show that $\lim_{n\to\infty} M_{Z_n}(t) = M_Z(t)$ for all $t \in (-h, h)$ for some $h > 0$.
Notice $Z \sim N(0,1)$ has mgf $M_Z(t) = e^{\frac{t^2}{2}}$, $t \in \mathbb{R}$ i.e. we need to show $\lim_{n\to\infty} M_{Z_n}(t) = e^{\frac{t^2}{2}}$ $\forall t \in (-h, h)$.
Note that $M_{Z_n}(t) = E[e^{tZ_n}] = E[e^{t\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}]$.
Notice that

$$\begin{aligned}
\frac{\bar{X}-\mu}{\sigma} &= \frac{\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \mu}{\sigma} \\
&= \frac{\frac{1}{n}\sum_{i=1}^n (\bar{X}-\mu)}{\sigma} \\
&= \frac{1}{n}\sum_{i=1}^n \frac{\bar{X}-\mu}{\sigma} \\
&= \frac{1}{n}\sum_{i=1}^n Y_i
\end{aligned}$$

where $Y_i = \frac{X_i - \mu}{\sigma}$, $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, therefore $E(Y_i) = 0$ and $Var(Y_i) = E[Y_i^2] = 1$. Y_is and X_is are both iid.

$$\begin{aligned}
M_{Z_n}(t) &= E[e^{t\sqrt{n}(\frac{\bar{X}-\mu}{\sigma})}] \\
&= E[e^{t\sqrt{n}\frac{1}{n}\sum_{i=1}^n Y_i}] \\
&= E[e^{\frac{t}{\sqrt{n}}(Y_1+\ldots+Y_n)}] \\
&= \big(M_{Y_i}(\frac{t}{\sqrt{n}})\big)^n
\end{aligned}$$

In summary $M_{Z_n}(t) = \big(M_{Y_i}(\frac{t}{\sqrt{n}})\big)^n$ where $E(Y_i) = 0$ and $Var(Y_i) = E(Y_i^2) = 1$.
Recall the Taylor expansion of $f(x)$ at a

$$f(x) = 1 + f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \ldots$$

Taking the Taylor expansion of $M_{Y_i}(\frac{t}{\sqrt{n}})$ about 0

$$
\begin{aligned}
M_{Y_i}(\frac{t}{\sqrt{n}}) &= M_{Y_i}(0) + \frac{M'_{Y_i}(0)}{1!}(\frac{t}{\sqrt{n}} - 0) + \frac{M''_{Y_i}(0)}{2!}(\frac{t}{\sqrt{n}})^2 + \frac{M^{(3)}_{Y_i}(0)}{3!}(\frac{t}{\sqrt{n}})^3 + \dots \\
&= E[e^{0Y_i}] + E[Y_i](\frac{t}{\sqrt{n}} - 0) + \frac{E[Y_i^2]}{2!}\frac{t^2}{n} + \frac{M^{(3)}_{Y_i}(0)}{3!}\frac{t^3}{\sqrt{n}}\frac{1}{n} + \frac{M^{(4)}_{Y_i}(0)}{4!}\frac{t^4}{n}\frac{1}{n} + \dots \\
&= 1 + \frac{t^2/2}{n} + \frac{1}{n}(\frac{M^{(3)}_{Y_i}(0)}{3!}\frac{t^3}{\sqrt{n}} + \frac{M^{(4)}_{Y_i}(0)}{4!}\frac{t^4}{n} + \dots)
\end{aligned}
$$

let $\psi(n) = \frac{M^{(3)}_{Y_i}(0)}{3!}\frac{t^3}{\sqrt{n}} + \frac{M^{(4)}_{Y_i}(0)}{4!}\frac{t^4}{n} + \dots$. Note that $\lim_{n\to\infty}\psi(n) = 0$ since $M^{(n)}_{Y_i}(0)$ are constant and t is constant so as $n \to \infty$ the term $\to 0$.
Thus we have

$$M_{Y_i}(\frac{t}{\sqrt{n}}) = 1 + \frac{t^2/2}{n} + \frac{\phi(n)}{n}$$

Thus

$$
\begin{aligned}
M_{Z_n}(t) &= \big(M_{Y_i}(\frac{t}{\sqrt{n}})\big)^n \\
&= \big(1 + \frac{t^2/2}{n} + \frac{\phi(n)}{n}\big)^n
\end{aligned}
$$

From the e-limit identity we have

$$\lim_{n\to\infty}\big(1 + \frac{b}{n} + \frac{\phi}{n}\big)^n = e^b \qquad \lim_{n\to\infty}\phi(n) = 0$$

Thus

$$\lim_{n\to\infty} M_{Z_n}(t) = e^{\frac{t^2}{2}}$$

We now need to show this holds for some neighbourhood around 0 for t.
Note that we assumed X_n's have mgf for $|t| < h$ (for all $t \in (-h, h)$, some $h > 0$). Also note that

$$
\begin{aligned}
M_{Y_i}(\frac{t}{\sqrt{n}}) &= E(e^{\frac{t}{\sqrt{n}}\frac{X_i - \mu}{\sigma}}) \\
&= e^{-\frac{t\mu}{\sqrt{n}\sigma}} E(e^{\frac{t}{\sqrt{n}\sigma}X_i}) \\
&= e^{-\frac{t\mu}{\sqrt{n}\sigma}} M_{X_i}(\frac{t}{\sqrt{n}\sigma})
\end{aligned}
$$

Therefore $M_{Y_i}(\frac{t}{\sqrt{n}})$ exists if $M_{X_i}(\frac{t}{\sqrt{n}\sigma})$ exists, i.e. $|\frac{t}{\sqrt{n}\sigma}| < h$ i.e. $|t| < \sqrt{n}\sigma h$ (similarly $M_{Z_n}(t)$ exists for the same t values).
Therefore $\lim_{n\to\infty} M_{Z_n}(t) = e^{\frac{t^2}{2}}$ for all $t \in (-h\sigma, h\sigma)$ then $Z_n \xrightarrow{D} Z \sim N(0, 1)$. $\square$

23 November 5, 2018

23.1 Example of CLT

Example 23.1. Suppose the number of times you check your smartphone follows $POI(5)$. Since there are 125 students let X_i be the # of times the ith student checks their point, then $X_i \overset{iid}{\sim} POI(5)$, $i = 1, \ldots, 125$.
Find $P(\bar{X} \leq 5.5)$ approximately.

Solution. $X_i \overset{iid}{\sim} POI(5)$ and $E(X_i) = Var(X_i) = \sigma^2 = 5$.
According to CLT we have $Z_n = \sqrt{n}\frac{(\bar{X}-\mu)}{\sigma} \xrightarrow{D} Z \sim N(0,1)$ that is $Z_n = \sqrt{n}\frac{(\bar{X}-5)}{\sqrt{5}} \xrightarrow{D} Z \sim N(0,1)$.
So

$$\begin{aligned} P(\bar{X} \leq 5.5) &= P(\sqrt{n}\frac{(\bar{X}-5)}{\sqrt{5}} \leq \sqrt{n}\frac{(5.5-5)}{\sqrt{5}}) \\ &= P(Z_n \leq \sqrt{n}\frac{0.5}{\sqrt{5}}) \end{aligned}$$

Notice $Z_n \xrightarrow{D} Z \sim N(0,1)$ therefore

$$\lim_{n\to\infty} P(Z_n \leq z) = P(Z \leq z) \qquad \forall z \in \mathbb{R}$$

Thus we have

$$P(\bar{X} \leq 5.5) = P(Z_n \leq \sqrt{125}\frac{0.5}{\sqrt{5}}) \approx P(Z_n \leq 2.5) = \Phi(2.5) = 0.9938$$

Example 23.2. Suppose $Y_n \sim \chi^2(n)$. Let $Z_n = \frac{Y_n - n}{\sqrt{2n}}$.
Show $Z_n \xrightarrow{D} Z \sim N(0,1)$.

Solution. Note that $Y_n \sim \chi^2(n) = \sum_{i=1}^n X_i$ where $X_i \overset{iid}{\sim} \chi^2(1)$ where $E(Y_i) = 1$ and $Var(Y_i) = 2 < \infty$.
According to CLT we have

$$\sqrt{n}\frac{(\bar{X}-1)}{\sqrt{2}} \xrightarrow{D} Z \sim N(0,1)$$

so

$$\sqrt{n}\frac{(Y_n/n-1)}{\sqrt{2}} = \frac{Y_n - n}{\sqrt{2n}} \xrightarrow{D} Z \sim N(0,1)$$

23.2 Summary of limiting behaviour

Convergence in distribution

1. $\lim_{n\to\infty} F_n(x) = F(x)$ for x where $F(\cdot)$ is continuous
2. $X_n \xrightarrow{P} X$ implies $X_n \xrightarrow{D} X$
3. MGF limit theorem
4. CLT

Convergence in probability

1. $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$
2. $X_n \xrightarrow{D} X = c$ iff $X_n \xrightarrow{P} X = c$
3. Weak law of large numbers (WLLN). Statement and proof below.

24 November 7, 2018

24.1 Weak law of large numbers (WLLN)

Theorem 24.1 (Weak law of large numbers (WLLN))**.** Suppose X_n's are *independent* (not necessarily identically distributed) r.v's and $E(X_n) = \mu$ and $Var(X_n) = \sigma^2 < \infty$, then $\bar{X} \xrightarrow{P} X = \mu$.

Remark 24.1. As long as n is large, we can use $\bar{X}$ to estimate μ as $\bar{X}$ is getting close to μ with probability approaching 1.

Proof. We will prove this using Markov's inequality.
To show $\bar{X} \xrightarrow{P} \mu$ we need to show

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$$

Notice $X_i \sim$ independent with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. So

$$E(\bar{X}) = E(\frac{1}{n}\sum_{i=1}^{\infty} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu$$

Also

$$Var(\bar{X}) = Var(\frac{1}{n}\sum_{i=1}^{\infty} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$$

For all $\epsilon > 0$ however small from Markov's inequality

$$\begin{aligned} P(|\bar{X} - \mu| > \epsilon) &\leq \frac{E(|\bar{X} - \mu|^2}{\epsilon^2} \\ &= \frac{E((\bar{X} - \mu)^2)}{\epsilon^2} \\ &= \frac{E((\bar{X} - E(\bar{X}))^2)}{\epsilon^2} \\ &= \frac{Var(\bar{X})}{\epsilon^2} \\ &= \frac{\sigma^2}{n\epsilon^2} \end{aligned}$$

Therefore

$$0 \leq \lim_{n\to\infty} P(|\bar{X} - \mu| > \epsilon) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0$$

i.e. $\lim_{n\to\infty} P(|\bar{X} - \mu| > \epsilon) = 0$ so $\bar{X} \xrightarrow{P} \mu$. □

24.2 Convergence preserved under certain transformations

The motivational example:

Example 24.1. Suppose $X_i \overset{iid}{\sim} POI(\mu)$, $\bar{X} \xrightarrow{P} \mu$ due to WLLN. Therefore we can use $\bar{X}$ to estimate μ.

Question 24.1. How to estimate $P(X_1 \leq 1)$?

Note that

$$\begin{aligned} P(X_1 \leq 1) &= f(0) + f(1) \\ &= \frac{e^{-\mu}\mu^0}{0!} + \frac{e^{-\mu}\mu^1}{1!} \\ &= e^{-\mu}(1+\mu) \end{aligned}$$

where we have a transformation of μ.
Recall that $\bar{X} \xrightarrow{P} \mu$ where we can estimate μ with $\bar{X}$. Does $e^{-\bar{X}}(1+\bar{X}) \xrightarrow{P} e^{-\mu}(1+\mu)$?
That is can the convergence in probability of $\bar{X}$ be preserved under this transformation?

Theorem 24.2 (Probability limit theorems). **(P1)** If $X_n \xrightarrow{P} a$ and $g(x)$ is continuous at $x = a$, then $g(X_n) \xrightarrow{P} g(a)$.

(P2) If $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$ and $g(x, y)$ is continuous at $x = a, y = b$, then $g(X_n, Y_n) \xrightarrow{P} g(a, b)$.

Theorem 24.3 (Distribution limit theorems). **(D1)** If $X_n \xrightarrow{D} X$ and $g(x)$ is continuous **on the entire support set of** X, then $g(X_n) \xrightarrow{D} g(X)$.

(D2) If $X_n \xrightarrow{D} X, Y_n \xrightarrow{P} b$ and $g(x, b)$ is continuous on the entire support set of X, then $g(X_n, Y_n) \xrightarrow{D} g(X, b)$.

This is also called the **Slutsky's theorem**.

Remark 24.2. In general if $X_n \xrightarrow{D} X, Y_n \xrightarrow{D} Y$, then $g(X_n, Y_n) \overset{D}{\not\to} g(X, Y)$.

Example 24.2. Suppose $X \sim N(0, 1)$ and let $X_n = X$ and $Y_n = -X$, $n = 1, 2, \ldots$.
Note that $X_N \sim N(0, 1)$ and $Y_n \sim N(0, 1)$ so $X_n, Y_n \xrightarrow{D} Z$.
Given $g(X_n, Y_n) = X_n + Y_n = X + (-X) = 0$ and clearly $0 \overset{D}{\not\to}= g(Z, Z) = 2Z \sim N(0, 4)$.

24.3 Example of limit probabilities/distributions of transformations

Example 24.3. Suppose $X_n \xrightarrow{P} a > 0$, $Y_n \xrightarrow{P} b > 0$, and $Z_n \xrightarrow{D} Z \sim N(0, 1)$. Show the limit probability/distribution of

1. X_n^2
2. $\sqrt{X_n}$
3. $X_n Y_n$
4. $X_n + Y_n$
5. $\frac{X_n}{Y_n}$
6. $2Z_n$
7. $Z_n + Y_n$
8. $X_n Z_n$
9. Z_n^2

We notice we can apply theorem (P1) to 1) and 2), (P2) to 3), 4), and 5), (P3) to 6) and 9), and (P4) to 7) and 8).

Solution.

1. Due to (P1), $X_n^2 \xrightarrow{P} a^2$ since $g(X_n) = X_n^2$ is continuous for any $a > 0$.
2. Due to (P1), $\sqrt{X_n} \xrightarrow{P} \sqrt{a}$ since $g(X_n) = \sqrt{X_n}$ is continuous for any $a > 0$ (if $a = 0$ then this would not hold).
3. Due to (P2), $X_nY_n \xrightarrow{P} ab$ since $g(X_n) = X_nY_n$ is continuous for any (a, b).
4. Similar as above by (P2), $X_n + Y_n \xrightarrow{P} a + b$.
5. Similar as above by (P2), $\frac{X_n}{Y_n} \xrightarrow{P} \frac{a}{b}$, $b \neq 0$.
6. Due to (D1), $2Z_n \xrightarrow{D} 2Z$ since $g(Z_n) = 2Z_n$ is continuous on $\mathbb{R}$ (which is the support set of Z).
7. Due to (D1), $Z_n^2 \xrightarrow{D} Z^2 \sim \chi^2(1)$ since $g(Z_n) = Z_n^2$ is continuous on $\mathbb{R}$ (which is the support set of Z).
8. Due to (D2), $Z_n + Y_n \xrightarrow{D} Z + b \sim N(b, 1)$ since $g(Z_n, Y_n) = Z_n + Y_n$ and $g(\cdot, b)$ is continuous on $\mathbb{R}$ (which is the support set of Z).
9. Similary as above by (D2), $X_nZ_n \xrightarrow{D} aZ \sim N(0, a^2)$.

25 November 9, 2018

25.1 Mean of Poisson r.v. and $P(X_1 \leq 1)$

Recall our motivational example with $X_i \overset{iid}{\sim} POI(\mu)$ with $\bar{X} \xrightarrow{P} \mu$ due to WLLN.

Question 25.1. Does $e^{-\bar{X}}(1 + \bar{X}) \xrightarrow{P} P(X_1 \leq 1) = e^{-\mu}(1 + \mu)$?
Yes. Let $g(x) = e^{-x}(1 + x)$ is continuous at $\mu \in \mathbb{R}$ therefore $g(\bar{X}) = e^{-\bar{X}}(1 + \bar{X}) \xrightarrow{P} g(\mu) = e^{-\mu}(1 + \mu)$ due to (P1).

25.2 More limit distribution examples

Example 25.1. Suppose $X_i \overset{iid}{\sim} POI(\mu)$. Let $Z_n = \frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{\bar{X}}}$. Show Z_n convergent in distribution and identify the limit distribution.

Solution. Since $X_i \overset{iid}{\sim} POI(\mu)$ with $E(X_i) = \mu$ and $Var(X_i) = \mu < \infty$ then $Y_n = \frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{\mu}} \xrightarrow{D} Z \sim N(0, 1)$ due to the CLT.
$\bar{X} \xrightarrow{P} \mu$ due to WLLN.
Note that

$$\begin{aligned} Z_n &= \frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{\bar{X}}} \\ &= \frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{\mu}} \frac{\sqrt{\mu}}{\sqrt{\bar{X}}} \\ &= Y_n \cdot Q_n \end{aligned}$$

where $Q_n = \frac{\sqrt{\mu}}{\sqrt{\bar{X}}} \xrightarrow{P} 1$ due to (P1) (where we have $g(x) = \frac{\sqrt{\mu}}{\sqrt{x}}$ is continuous at $x = \mu > 0$).
Due to Slutsky's theorem (D2) we have

$$Z_n = Y_n \cdot Q_n \xrightarrow{D} Z \cdot 1 = Z \sim N(0, 1)$$

Remark 25.1. Why is $\frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{\bar{X}}} \xrightarrow{D} Z \sim N(0,1)$ useful?
In comparison: $Y_n = \frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{\mu}} \xrightarrow{D} Z \sim N(0,1)$ is NOT always directly useful in practice. For example when constructing confidence interval for μ we can utilize

$$Z_n = \frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{\bar{X}}} \xrightarrow{D} Z \sim N(0,1)$$

then $P(Z_{0.05} \leq Z_n \leq Z_{0.95}) \approx 0.9$ i.e. $P(-1.645 \leq \frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{\bar{X}}} \leq 1.645) \approx 0.9$, which we can use to find an approximate 90% CI for μ

$$\begin{aligned} &-1.645 \leq \frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{\bar{X}}} \leq 1.645 \\ \Rightarrow &\bar{X} - 1.645\sqrt{\frac{\bar{X}}{n}} \leq \mu \leq \bar{X} + 1.645\sqrt{\frac{\bar{X}}{n}} \end{aligned}$$

which we can estimate with $\bar{X}$. If we used Y_n instead our confidence interval would be in terms of μ which is we do not know (since we are estimating μ).

25.3 Delta method

The Delta method states preservation of convergence in distribution to the normal distribution.

Theorem 25.1 (Delta method)**.** Suppose $Y_1, \ldots, Y_n$ are a sequence of r.v.s such that $\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \sigma^2)$ and $g(\cdot)$ is differentiable at θ and $g'(\theta) \neq 0$.
Then $\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{D} N(0, (g'(\theta))^2\sigma^2)$.

Proof. Taylor expansion of $g(Y_n)$ at θ

$$\begin{aligned} &g(Y_n) = g(\theta) + g'(\theta)(Y_n - \theta) + \ldots \\ \Rightarrow &\sqrt{n}(g(Y_n) - g(\theta)) = g'(\theta)\sqrt{n}(Y_n - \theta) \end{aligned}$$

where $g'(\theta)$ is constant and $\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \sigma^2)$. Therefore

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{D} g'(\theta)N(0, \sigma^2)$$

due to (D1), where $g'(\theta)N(0, \sigma^2) = N(0, g'(\theta)^2\sigma^2)$. □

Example 25.2. Let $X_i \overset{iid}{\sim} EXP(\theta)$, $i = 1, \ldots, n$. Find the limit distribution of

1. $\bar{X}$
2. $Z_n = \frac{\sqrt{n}(\bar{X}-\theta)}{\bar{X}}$
3. $U_n = \sqrt{n}(\bar{X} - \theta)$
4. $V_n = \sqrt{n}(\log \bar{X} - \log \theta)$

Solution.
1. Exercise.
2. Exercise.

3. Since $E(X_i) = \theta$ and $Var(X_i) = \theta^2 < \infty$ then by the CLT $Y_n = \frac{\sqrt{n}(\bar{X}-\theta)}{\theta} \xrightarrow{D} \sim N(0,1)$.

   So $U_n = \theta Y_n \xrightarrow{D} \theta \cdot Z \sim N(0, \theta^2)$ due to (D1).

4. Note that $\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \theta^2)$ from (3).

   Note that $g(x) = \log(x)$ is differentiable at $\theta > 0$ ($g'(x) = \frac{1}{x}$) and $g'(\theta) = \frac{1}{\theta} \neq 0$.

   By the Delta method we have

   $$V_n = \sqrt{n}(\log \bar{X} - \log \theta) \xrightarrow{D} N(0, (g'(\theta))^2\theta^2) = N(0, (\frac{1}{\theta})^2\theta^2)$$

   So $V_n \xrightarrow{D} N(0,1)$.

   Remark 25.2. We can use V_n to construct CI's for $\log \theta$.

25.4 Estimation (point estimation and set estimation)

Suppose $X_1, \ldots, X_n$ are iid r.v.s with pf/pdf $f(x;\theta)$ (paramterized by parameter(s) θ). θ is an unknown parameter and $\theta \in \Omega$ which contains all possible values of θ (i.e. the **parameter space**).

Example 25.3. If $X_i \overset{iid}{\sim} N(\mu, 1)$, we have one parameter $\theta = \mu$ and $\Omega = \mathbb{R}$.

Example 25.4. If $X_i \overset{iid}{\sim} GAM(\alpha, \beta)$, we have two parameters $th\vec{e}ta = (\alpha, \beta)$ and $\Omega = \{(\alpha,\beta) \mid \alpha, \beta > 0\} = \mathbb{R}^+ \times \mathbb{R}^+$.

We are interested in **estimating** θ based on a random sample $(x_1, \ldots, x_n)$. There are two main topics:

Point estimation A **single** possible value for θ

Set estimation A set of possible values for θ (i.e. confidence interval)

26 November 12, 2018

26.1 Estimators and estimates

For a given parameter θ the information about θ is hidden in the random sample. To estimate θ is to extract such information.

Definition 26.1 (Statistic)**.** A **statistic** $T = T(X_1, \ldots, X_n)$ is a function of the random sample which does NOT depend on any unknown parameter(s).

Example 26.1. Sample mean

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = T(X_1, \ldots, X_n)$$

is a statistic.

Sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = T(X_1, \ldots, X_n)$$

is a statistic.

Note that $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ is NOT a statistic since it depends on unknown parameters μ and σ.

Definition 26.2 (Estimator)**.** A statistic $T = T(X_1, \ldots, X_n)$ used to estimate θ is called an **estimator** of θ.

Definition 26.3 (Estimate)**.** An observed value of the statistic T i.e. $t = T(x_1, \ldots, x_n)$ is called an **estimate** of θ.

Remark 26.1. An estimator is a random variable but an estimate is NOT.

Example 26.2. $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ is an *estimator* of μ whereas $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ is an *estimate* of μ.

26.2 Maximum likelihood estimation

Definition 26.4 (Likelihood function)**.** Suppose $X_i \overset{iid}{\sim} f(x;\theta)$ (the pf/pdf), and given the observed sample $(x_1, \ldots, x_n)$ the **likelihood function** $L(\theta)$ is defined as the joint pf/pdf of $(x_1, \ldots, x_n)$ i.e.

$$L(\theta; x) = \prod_{i=1}^n f(x_i;\theta) = L(\theta)$$

Remark 26.2. 1. Discrete case We have

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n f(x_i;\theta) \\ &= \prod_{i=1}^n P(X_i = x_i;\theta) \\ &= P(X_1 = x_1, \ldots, X_n = x_n;\theta) && \text{independence} \end{aligned}$$

Therefore $L(\theta)$ is the probability of the observed $(x_1, \ldots, x_n)$ with a given θ.

2. Comparison of $L(\theta)$ If $L(\theta_1) < L(\theta_2)$ where $\theta_1, \theta_2 \in \Omega$ and for the same set of observations $(x_1, \ldots, x_n)$ i.e.

$$P(X_1 = x_1, \ldots, X_n = x_n;\theta_1) = L(\theta_1) < L(\theta_2) = P(X_1 = x_1, \ldots, X_n = x_n;\theta_2)$$

This means that it is more likely to obtain our observed $(x_1, \ldots, x_n)$ with $\theta = \theta_2$ than $\theta = \theta_1$.

3. MLE The idea of the maximum likelihood estimation is to find the value(s) of θ that can maximize $L(\theta)$.

Definition 26.5 (Maximum Likelihood Estimate (MLE))**.** The **Maximum Likelihood Estimate (MLE)** of θ is

$$\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n) = \text{argmax}_{\theta\in\Omega} L(\theta)$$

Remark 26.3. $\hat{\theta}$ is the value of θ within Ω such that $L(\theta)$ is maximized (i.e. $\hat{\theta}$ is a maximizer of $L(\theta)$), that is $L(\hat{\theta}) \geq L(\theta)$ for all $\theta \in \Omega$.

Definition 26.6 (Maximum likelihood estimator)**.** The maximum likelihood estimator of θ is $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ (function of our random variable samples X_i).

Example 26.3. If $X_i \overset{iid}{\sim} POI(\theta)$, $i = 1, \ldots, n$ and $\theta > 0$ (i.e. $\Omega = \mathbb{R}^+$), find the MLE of θ.

Solution. Step 1: Find $L(\theta)$ Our likelihood function is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\ &= \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} \\ &= \frac{e^{-n\theta}\theta^{\sum x_i}}{\prod x_i!} \end{aligned}$$

Step 2: Find $L'(\theta)$ We can either find $L'(\theta) = \frac{\partial L(\theta)}{\partial \theta} = 0$ to solve for $\hat{\theta}$ or instead solve for $l'(\theta) = \frac{\partial \log L(\theta)}{\partial \theta} = 0$ (the partial derivative of the log likelihood function).

Remark 26.4. Since the log transformation is strictly montonically increasing then $\hat{\theta}$ maximizing $\log L(\theta)$ is equivalent to maximizing $L(\theta)$.

Taking the derivation of $\log L(\theta)$ is easier since $L(\theta)$ is a product of θ terms whereas $\log L(\theta)$ is a summation of θ terms.

So we have

$$\begin{aligned} l(\theta) &= \log e^{-n\theta} + \log \theta^{\sum x_i} - \log \prod x_i! \\ &= -n\theta + \sum_{n=1}^{\infty} x_i \log\theta - \sum_{n=1}^{\infty} \log x_i! \end{aligned}$$

Thus we have

$$l'(\theta) = -n + \frac{\sum_{i=1}^{\infty} x_i}{\theta}$$

Solving for $l'(\theta) = 0$ to find $\hat{\theta}$ we get that

$$-n + \frac{\sum_{i=1}^{n} x_i}{\theta} = 0 \Rightarrow \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Step 3: Verify $\hat{\theta} = \bar{x}$ is maximizer of $l(\theta)$ We can use either the 1st OR 2nd derivative test:

1st deriative test Check that $l'(\theta) > 0$ if $\theta < \hat{\theta}$ and $l'(\theta) < 0$ if $\theta > \hat{\theta}$

2nd deriative test Check that

$$\left.\frac{\partial^2 l(\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}}$$

i.e. $l''(\hat{\theta}) < 0$.

So back to our example we have

1st deriative test

$$
\begin{aligned}
l'(\theta) &= -n + \frac{\sum_{i=1}^{n} x_i}{\theta} \\
&= -n + \frac{n\hat{\theta}}{\theta} \\
&= (\frac{\hat{\theta}}{\theta} - 1)n
\end{aligned}
$$

If $\theta < \hat{\theta}$ then $\frac{\hat{\theta}}{\theta} > 1$ therefore $l'(\theta) > 0$. Similarly if $\theta > \hat{\theta}$ then clearly $l'(\theta) < 0$.

2nd deriative test

$$
l''(\theta) = -\frac{\sum_{i=1}^{n} x_i}{\theta^2}
$$

Clearly $l''(\theta)$ for all θ.

Therefore $\hat{\theta} = \bar{x}$ is the MLE of θ.

Exercise 26.1. If $X \overset{iid}{\sim} EXP(\theta)$, find the MLE of θ.

Remark 26.5. What if $\frac{\partial l(\theta)}{\partial \theta} = 0$ has more than one solution? Suppose we have $\hat{\theta}_1$ and $\hat{\theta}_2$. Verify that they are maximizers and then compare $L(\hat{\theta}_1)$ with $L(\hat{\theta}_2)$. If one of $L(\hat{\theta}_i)$ is bigger then $\hat{\theta}_i$ is the MLE. If they are equal then they are both MLEs of θ.

27 November 14, 2018

27.1 Bernoulli MLE

Suppose we conduct n iid $BERN(\theta)$ trials and we observe x successes. Find the MLE of θ.

Solution. Note that $X \sim BIN(n, \theta)$, $X = x$ is observed. We use the Binomial distribution with 1 sample (instead of n Bernoulli samples) for our likelihood function.

Step 1: $l(\theta)$ Note that

$$
\begin{aligned}
L(\theta) &= f(x; \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \\
l(\theta) &= \log \binom{n}{x} x \log \theta + (n-x) \log(1-\theta)
\end{aligned}
$$

Step 2: $\frac{\partial(\theta)}{\partial \theta}$

$$
\frac{\partial l(\theta)}{\partial \theta} = \frac{x}{\theta} + \frac{n-x}{1-\theta}(-1)
$$

Solving for $\frac{\partial l(\theta)}{\partial \theta} = 0$ we get $\hat{\theta} = \frac{x}{n}$.

Step 3: first derivative test Note that

$$\begin{aligned} l'(\theta) &= \frac{x(1-\theta)-\theta(n-x)}{\theta(1-\theta)} \\ &= \frac{x-n\theta}{\theta(1-\theta)} \\ &= \frac{n(\frac{x}{n}-\theta)}{\theta(1-\theta)} \\ &= \frac{n(\hat{\theta}-\theta)}{\theta(1-\theta)} \end{aligned}$$

Note $\theta(1-\theta) > 0$ as $0 < \theta < 1$.

Therefore $l'(\theta) > 0$ if $n(\hat{\theta}-\theta) > 0$ i.e. $\theta < \hat{\theta}$. Similarly $l'(\theta) < 0$ if $n(\hat{\theta}-\theta) < 0$ i.e. $\theta > \hat{\theta}$.

So $\hat{\theta}$ is indeed a (global) maximum of the log-likelihood function.

Exercise 27.1. Let $X_i \overset{iid}{\sim} EXP(\theta)$, then $\hat{\theta} = \bar{x}$.

27.2 Information functions

To prepare for the limit distribution of the MLE, we introduce the following **information functions** $S(\theta), I(\hat{\theta}), I(\theta)$.

Definition 27.1 (Score function)**.** The **score function** $S(\theta)$ is defined as

$$S(\theta) = \frac{\partial l(\theta)}{\partial \theta}$$

Note that we solve $S(\theta) = 0$ to find the MLE $\hat{\theta}$.

Definition 27.2 (Information function)**.** The **information function** $I(\theta)$ is defined as

$$I(\theta) = -\frac{\partial^2 l(\theta)}{\partial \theta^2} = -\frac{\partial S(\theta)}{\partial \theta}$$

Definition 27.3 (Observed information function)**.** The **observed information function** $I(\hat{\theta})$ is defined as

$$I(\hat{\theta}) = I(\theta)\Big|_{\theta=\hat{\theta}}$$

Example 27.1. Returning to our previous example where $X_i \overset{iid}{\sim} POI(\theta)$, $\hat{\theta} = \bar{x}$.
Note $S(\theta) = \frac{\partial l(\theta)}{\partial \theta} = -n + \frac{\sum_{i=1}^n X_i}{\theta}$ then

$$I(\theta) = -\Big(-\frac{\sum_{i=1}^n X_i}{\theta^2}\Big) = \frac{\sum_{i=1}^n X_i}{\theta^2}$$

and

$$\begin{aligned} I(\hat{\theta}) &= \frac{\sum_{i=1}^n x_i}{\hat{\theta}^2} \\ &= \frac{\sum_{i=1}^n x_i}{\bar{x}^2} \\ &= \frac{n}{\bar{x}} \end{aligned}$$

Remark 27.1. If the sample size increases you should have more information about the unknown parameter θ, and the estimate should be more trustworthy. This is reflected in $I(\hat{\theta}) = \frac{n}{\bar{x}}$ which increases as sample size increases.

Remark 27.2. Recall $L(\theta)$ is not only a function of θ, but also a function of $\boldsymbol{x} = (x_1, \ldots, x_n)$ as $L(\theta) = \prod_{i=1}^n f(X_i; \theta)$.
Therefore $l(\theta), S(\theta)$ and $I(\theta)$ are all also functions of $\boldsymbol{x}$.
When we are interested in the properties of these functions of the random sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ we use $L(\theta; \boldsymbol{X})$, $l(\theta; \boldsymbol{X})$, etc.

27.3 Fisher (or expected) information function

Definition 27.4 (Fisher information function)**.** The **Fisher information function** is defined as

$$J(\theta) = E[I(\theta; \boldsymbol{X})]$$

that is the expectation of our information function **over our random sample**.

Example 27.2. Returning again to our previous $X_i \overset{iid}{\sim} POI(\theta)$ example, recall $I(\theta) = \frac{\sum_{i=1}^n x_i}{\theta^2} = I(\theta; \boldsymbol{x})$. Then

$$\begin{aligned} J(\theta) &= E[I(\theta; \boldsymbol{x})] \\ &= E\big[\frac{\sum_{i=1}^n X_i}{\theta^2}\big] \\ &= \frac{\sum_{i=1}^n E[X_i]}{\theta^2} \\ &= \frac{n\theta}{\theta^2} \\ &= \frac{n}{\theta} \end{aligned}$$

Recall $\hat{\theta} = \bar{X}$ and $Var(\bar{X}) = \frac{\theta}{n} = \frac{1}{J(\theta)}$.

Remark 27.3. The large the value of $J(\theta)$ (i.e. more information) the smaller the value of $Var(\hat{\theta})$ i.e. $\hat{\theta}$ is a more precise estimator.

27.4 MLE when support set depends on θ

What if the support set depends on the unknown parameter θ and we need to find the MLE of θ?

Example 27.3. Suppose $X_i \overset{iid}{\sim} UNIF(0, \theta)$, $\theta > 0$. Note that $f(x; \theta) = \frac{1}{\theta}$ if $0 \leq x \leq \theta$.
Thus

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) = \big(\frac{1}{\theta}\big)^n$$

if $0 \leq x_i \leq \theta$, $i = 1, \dots, n$.
Note that $L(\theta) = \frac{1}{\theta^n}$ is strictly decreasing for $\theta > 0$.
To maximize $L(\theta)$, we need to essentially minimize θ. However, if $\theta < x_i$ for some x_i, then $L(\theta) = 0$ since $f(x_i; \theta) = 0$ for $x_i > \theta$.
Therefore the smallest θ that maintains $L(\theta) = \frac{1}{\theta^n} \neq 0$ is $\hat{\theta} = \max(x_1, \dots, x_n)$ which is the MLE of θ.
What if we tried to use our initial method solving for $S(\theta) = 0$? Note that $L(\theta) = \frac{1}{\theta^n}$ (if $0 \leq x_i \leq \theta$ for $i = 1, \dots, n$) thus $l(\theta) = -n \log \theta$ and $S(\theta) = \frac{\partial l(\theta)}{\partial \theta} = \frac{-n}{\theta}$.
We cannot simply solve for $S(\theta) = 0$ to find $\hat{\theta}$ since we have a restriction the valid x_i values for $L(\theta)$.
So actually, $L(\theta) = \frac{1}{\theta^n} \cdot I[0 \leq x_i \leq \theta, i = 1, \dots, n]$ where I is the indicator function.
Taking the derivative with respect to θ for an indicator function is troublesome (but not impossible). Instead we use the method above to find $\hat{\theta}$.

28 November 16, 2018

28.1 MLE example for non-parameter values

Example 28.1. Suppose $X_i \overset{iid}{\sim} POI(\theta)$, $i = 1, \dots, n$, $\theta > 0$.
Find the MLE of $\tau = P(X_1 = 0)$.

Solution. To find the MLE of θ, we found $L(\theta)$, $l(\theta)$, $S(\theta)$ and solved $S(\theta) = 0$ to find $\hat{\theta}$.
Note that $\tau = P(X_1 = 0) = \frac{e^{-\theta}\theta^0}{0!} = e^{-\theta}$.
To find the MLE of τ, we find $L(\tau)$, $l(\tau)$, $S(\tau)$ and solve $S(\tau) = 0$ to find $\hat{\tau}$.
Note that $L(\tau) = \prod_{i=1}^n f(x_i; \tau)$.
Recall $f(x; \theta) = \frac{e^{-\theta}\theta^x}{x!}$ and $\tau = e^{-\theta}$, $\theta = -\log \tau$ so $f(x; \tau) = \frac{\tau(-\log \tau)^x}{x!}$.
Therefore

$$\begin{aligned} L(\tau) &= \prod_{i=1}^n \frac{\tau(-\log \tau)^{x_i}}{x_i!} \\ &= \frac{\tau^n(-\log \tau)^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \end{aligned}$$

and

$$l(\tau) = n \log \tau + \sum_{i=1}^n x_i \log(-\log \tau) - \log \prod_{i=1}^n x_i!$$

so we have

$$\begin{aligned} S(\tau) &= \frac{\partial l(\tau)}{\partial \tau} \\ &= \frac{n}{\tau} + \frac{\sum_{i=1}^n x_i}{-\log \tau}(-\frac{1}{\tau}) \\ &= \frac{n \log \tau + \sum_{i=1}^n x_i}{\tau(\log \tau)} \end{aligned}$$

Solving $S(\tau) = 0$ to find $n \log \tau + \sum_{i=1}^n x_i = 0$ so $\hat{\tau} = e^{-\bar{x}}$.
We now verify $S(\tau) > 0$ for $\tau < \hat{\tau}$ and $S(\tau) < 0$ if $\tau > \hat{\tau}$.
Since $S(\tau) = \frac{n \log \tau + \sum_{i=1}^n x_i}{\tau(\log \tau)}$, note that $\tau \log \tau < 0$ as $\tau = P(X_1 = 0) \in (0, 1)$.
$S(\tau) > 0$ if $n \log \tau + \sum_{i=1}^n x_i < 0$ (i.e. $\tau < e^{-\bar{x}} = \hat{\tau}$). Similarly $S(\tau < 0)$ if $\tau > \hat{\tau}$.

Thus the MLE of τ is $\hat{\tau} = e^{-\bar{x}}$.

28.2 Invariance property of MLE

Theorem 28.1 (Invariance property of MLE)**.** If $\hat{\theta}$ is the MLE of θ and $\tau = \tau(\theta)$ is a 1-to-1 function of θ then the MLE of τ is $\hat{\tau} = \tau(\hat{\theta})$.

Note. In our previous example we had $\hat{\theta} = \bar{x}$ and $\tau = e^{-\theta} = \tau(\theta)$. τ is 1-to-1 since $\tau(\theta)$ is strictly decreasing in terms of θ).
Then according to our theorem, the MLE of τ is $\tau(\hat{\theta}) = e^{-\hat{\theta}} = e^{-\bar{x}}$.

Example 28.2. Suppose $X_i \overset{iid}{\sim} f(x;\theta) = \theta x^{\theta-1}$, for $0 < x < 1$, $\theta > 0$. Find the MLE of the median of the distribution $f(x;\theta)$.
From a previous result we found that $\hat{\theta} = -\frac{n}{\sum_{i=1}^n \log x_i}$. Let τ be the median of $f(x;\theta)$. We need to find $\tau = \tau(\theta)$ a 1-to-1 function of θ to apply our theorem such that $\hat{\theta} = \tau(\hat{\theta})$.
Let $F(\cdot)$ be the cdf of $f(x;\theta)$ where $F(\tau) = P(X \leq \tau) = \frac{1}{2}$.
Since $f(x;\theta) = \theta x^{\theta-1}$ for $0 < x < 1$ then

$$F(x) = \int_0^x f(x;\theta)\,\mathrm{d}x = \int_0^x \theta x^{\theta-1}\,\mathrm{d}x = x^\theta$$

Thus $F(\tau) = \tau^\theta = \frac{1}{2}$ for $0 < \tau < 1$ that is $\tau = \left(\frac{1}{2}\right)^{\frac{1}{\theta}} = \tau(\theta)$.
Note that $\tau(\theta)$ is strictly increasing in terms of θ, then $\hat{\tau} = \tau(\hat{\theta}) = \left(\frac{1}{2}\right)^{\frac{1}{\hat{\theta}}}$.

28.3 Newton's method

Example 28.3. Suppose $X_i \overset{iid}{\sim} WEI(1,\theta)$, $\theta > 0$, $i = 1, \ldots, n$.
Find the MLE of θ.

Solution.

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n f(x_i;\theta) \\ &= \prod_{i=1}^n \theta x_i^{\theta-1} e^{-x_i^\theta} \\ &= \theta\Big(\prod_{i=1}^n x_i\Big)^{\theta-1} e^{-\sum_{i=1}^n x_i^\theta} \end{aligned}$$

Note that

$$l(\theta) = n\log\theta(\theta-1)\sum_{i=1}^n \log x_i - \sum_{i=1}^n x_i^\theta$$

So we have

$$S(\theta) = \frac{\partial l(\theta)}{\partial\theta} = \frac{n}{\theta} + \sum_{i=1}^n \log x_i - \sum_{i=1}^n (\log x_i) x_i^\theta$$

We solve for $S(\theta) = 0$ i.e.

$$\frac{n}{\theta} = \sum_{i=1}^n (\log x_i) x_i^\theta - \sum_{i=1}^n \log x_i$$

Note that there is no analytical solution(s) for $S(\theta) = 0$.
We thus must find $\hat{\theta}$ numerically through iteration e.g. using **Newton's method**.

Step 1 Find the initial value $\theta^{(0)}$.

Step 2 Update $\theta^{(r+1)} = \theta^{(r)} + \frac{S(\theta^{(r)})}{I(\theta^{(r)})}$.

Step 3 Let $\theta^{(k)}$ be the numerical solution to $S(\theta) = 0$ if $|\theta^{(k+1)} - \theta^{(k)}| < \epsilon$, where $\epsilon > 0$ is a pre-specified small value e.g. $\epsilon = 10^{-5}$.

Remark 28.1. Verify $\theta^{(k)}$ is a maximizer at the end of step 3.

So for our example we have

$$S(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} (\log x_i) x_i^{\theta}$$

and

$$\begin{aligned} I(\theta) = -\frac{\partial S(\theta)}{\partial \theta} &= -\Big(-\frac{n}{\theta^2} - \sum_{i=1}^{n} (\log x_i)^2 x_i^{\theta} \Big) \\ &= \frac{n}{\theta^2} + \sum_{i=1}^{n} (\log x_i)^2 x_i^{\theta} \end{aligned}$$

Therefore

$$\theta^{(r+1)} = \theta^{(r)} + \frac{\frac{n}{\theta^{(r)}} + \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} (\log x_i) x_i^{\theta^{(r)}}}{\frac{n}{\theta^{(r)2}} + \sum_{i=1}^{n} (\log x_i)^2 x_i^{\theta^{(r)}}}$$

Remark 28.2. Newton's method does NOT always converge unless the "quadratic convergence conditions" are satisfied (not required for STAT 330).

Graphical intuition for why Newton's method works:

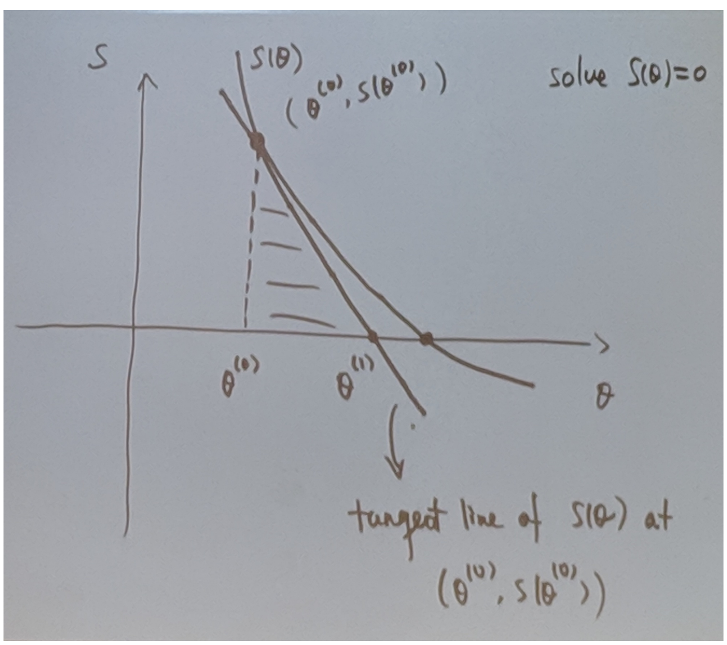


Figure 28.1: Graph of $S(\theta)$ with tangent draw at $(\theta^{(0)}, S(\theta^{(0)}))$.

If we take a step towards our root ($S(\theta) = 0$), we essentially walk along the tangent line at $\theta^{(0)}$.

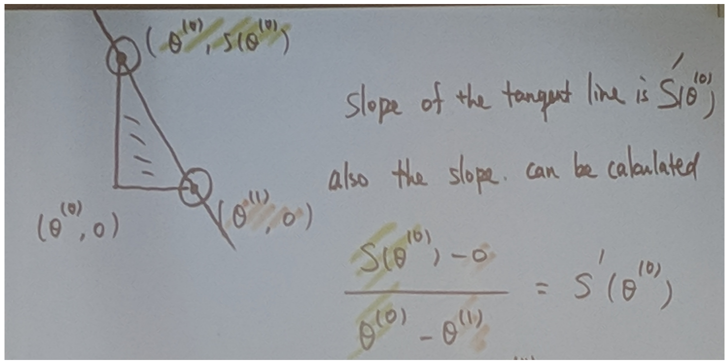


Figure 28.2: Tangent line where $S(\theta) = 0$ occurring at $\theta = \theta^{(1)}$.

Note that the tangent line to $S(\theta) = 0$ is given as

$$\begin{aligned}
\frac{S(\theta^{(0)}) - 0}{\theta^{(0)} - \theta^{(1)}} &= S'(\theta^{(0)}) \\
\Rightarrow \theta^{(1)} &= \theta^{(0)} + \frac{S(\theta^{(0)})}{-S'(\theta^{(0)})} \\
\Rightarrow \theta^{(1)} &= \theta^{(0)} + \frac{S(\theta^{(0)})}{I(\theta^{(0)})}
\end{aligned}$$

29 November 19, 2018

29.1 Asymptotic properties of ML estimators

We "verify" the four **asymptotic properties** for various distributions satisfying "regularity conditions" (discussed more in STAT 450):

1. $\hat{\theta} \xrightarrow{P} \theta$ (we verify this by apply WLLN or (P1)).
2. (Recall: $J(\theta) = E(I(\theta))$ the Fisher information function).

   $\sqrt{J(\theta)}(\hat{\theta} - \theta) \xrightarrow{D} Z \sim N(0, 1)$ (we verify this by applying CLT, (D1), (D2) i.e. Slutsky's, and/or the Delta method).
3. $\sqrt{J(\hat{\theta})}(\hat{\theta} - \theta) \xrightarrow{D} Z \sim N(0, 1)$ (we verify this by applying CLT, (D1), (D2) i.e. Slutsky's, and/or the Delta method).
4. $\sqrt{I(\hat{\theta})}(\hat{\theta} - \theta) \xrightarrow{D} Z \sim N(0, 1)$ (we verify this by applying CLT, (D1), (D2) i.e. Slutsky's, and/or the Delta method).

Example 29.1. Suppose $X_i \overset{iid}{\sim} f(x; \theta) = \theta x^{\theta - 1}$, $0 < x < 1$ and $\theta > 0$ ($\Omega = \mathbb{R}^+$). Find $\hat{\theta}$, $S(\theta)$, $I(\theta)$, $I(\hat{\theta})$, and $J(\theta)$ and verify our four asymptotic properties.

Solution. Step 1 Note that

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(X_i;\theta) \\ &= \prod_{i=1}^{n} \theta X_i^{\theta-1} \\ &= \theta^n \big(\prod_{i=1}^{n} X_i\big)^{\theta-1} \end{aligned}$$

so

$$l(\theta) = n\log\theta + (\theta-1)\sum_{i=1}^{n}\log X_i$$

Step 2 Note that

$$S(\theta) = \frac{\partial l(\theta)}{\partial\theta} = \frac{n}{\theta} + \sum_{i=1}^{n}\log X_i$$

solving $S(\theta) = 0$ we get $\hat\theta = -\frac{n}{\sum_{i=1}^{n}\log X_i}$.

Step 3 To show $\hat\theta$ is the MLE we use the 2nd derivative test.

Note that $S'(\theta) = -\frac{n}{\theta^2} < 0$ for any $\theta > 0$.

Therefore $S'(\hat\theta) < 0$ thus $\hat\theta$ is the MLE of θ.

Step 4 We have $I(\theta) = -\frac{\partial S(\theta)}{\partial\theta} = \frac{n}{\theta^2}$.

Also

$$J(\theta) = E(I(\theta \mid X)) = E(\frac{n}{\theta^2}) = \frac{n}{\theta^2}$$

and $I(\hat\theta) = \frac{n}{(\hat\theta)^2} = J(\hat\theta)$.

Step 5 We verify the asymptotic properties:

1. Verify $\hat\theta \xrightarrow{P} \theta$ where $\hat\theta = -\frac{n}{\sum_{i=1}^{n}\log X_i}$.

   Notice that

   $$\hat\theta = \frac{1}{\frac{\sum_{i=1}^{n} -\log X_i}{n}} = \frac{1}{\bar Y}$$

   where $Y_i = -\log X_i$.

   Since $X_i \overset{iid}{\sim} f(x;\theta) = \theta x^{\theta-1}$ for $0 < x < 1$ and $-\log(\cdot)$ is a 1-1 function, we have $Y_i \overset{iid}{\sim} g(y;\theta) = \theta e^{-y\theta}$ for $y > 0$.

   Therefore $Y_i \overset{iid}{\sim} GAM(1, \frac{1}{\theta})$ where $E(Y_i) = \frac{1}{\theta}$ and $Var(Y_i) = \frac{1}{\theta^2} < \infty$.

   By the WLLN we have $\bar Y \xrightarrow{P} E(Y_i) = \frac{1}{\theta}$.

   Due to (P1) let $g(x) = \frac{1}{x}$ where $g(\cdot)$ is continuous at $\frac{1}{\theta}$ for $\theta > 0$ then

   $$g(\bar Y) = \frac{1}{\bar Y} = \hat\theta \xrightarrow{P} \frac{1}{\frac{1}{\theta}} = \theta$$

2. Verify $\sqrt{J(\theta)}(\hat{\theta}-\theta) \xrightarrow{D} Z \sim N(0,1)$.
   Note that
$$\sqrt{J(\theta)}(\hat{\theta}-\theta) = \sqrt{\frac{n}{\theta^2}}\left(\frac{1}{\bar{Y}} - \frac{1}{\frac{1}{\theta}}\right)$$
   By the CLT we know that
$$\frac{\sqrt{n}(\bar{Y} - E(Y_1))}{\sqrt{Var(Y_1)}} \xrightarrow{D} Z \sim N(0,1)$$
   that is
$$\frac{\sqrt{n}(\bar{Y} - \frac{1}{\theta})}{\sqrt{\frac{1}{\theta^2}}} \xrightarrow{D} Z \sim N(0,1)$$
   we transform this using (D1)
$$\left(\frac{1}{\theta}\right)\frac{\sqrt{n}(\bar{Y} - \frac{1}{\theta})}{\frac{1}{\theta}} \xrightarrow{D} \left(\frac{1}{\theta}Z\right) \sim N\left(0, \frac{1}{\theta^2}\right)$$
   that is
$$\sqrt{n}\left(\bar{Y} - \frac{1}{\theta}\right) \xrightarrow{D} N\left(0, \frac{1}{\theta^2}\right)$$
   Recall by the Delta method if $\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \sigma^2)$ and $g'(\theta) \neq 0$ then $\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{D} N(0, \left(g'(\theta)\right)^2 \sigma^2)$.
   We let $g(x) = \frac{1}{x}$ where $g'(x) = -\frac{1}{x^2} \neq 0$. To apply the Delta method we require $g'(\frac{1}{\theta}) = -\theta^2 \neq 0$.
   Since $\sqrt{n}(\bar{Y} - \frac{1}{\theta}) \xrightarrow{D} N(0, \frac{1}{\theta^2})$ applying the Delta method we have
$$\sqrt{n}\left(\frac{1}{\bar{Y}} - \frac{1}{\frac{1}{\theta}}\right) \xrightarrow{D} N\left(0, \left(g'\left(\frac{1}{\theta}\right)\right)^2 \frac{1}{\theta^2}\right)$$
   note that $g'(\frac{1}{\theta}) = -\frac{1}{\frac{1}{\theta^2}} = -\theta^2$ thus
$$\sqrt{n}\left(\frac{1}{\bar{Y}} - \frac{1}{\frac{1}{\theta}}\right) \xrightarrow{D} N(0, \theta^2)$$
   that is
$$\left(\frac{1}{\theta}\right)\sqrt{n}\left(\frac{1}{\bar{Y}} - \frac{1}{\frac{1}{\theta}}\right) \xrightarrow{D} N(0,1)$$
   so we have
$$\sqrt{\frac{n}{\theta^2}}(\hat{\theta}-\theta) = \sqrt{J(\theta)}(\hat{\theta}-\theta) \xrightarrow{D} N(0,1)$$

3. Verify $\sqrt{J(\hat{\theta})}(\hat{\theta}-\theta) \xrightarrow{D} N(0,1)$.
   Note that $\sqrt{J(\hat{\theta})}(\hat{\theta}-\theta) = \sqrt{J(\theta)}(\hat{\theta}-\theta)\sqrt{\frac{J(\hat{\theta})}{J(\theta)}}$.
   We know $\sqrt{J(\theta)}(\hat{\theta}-\theta) \xrightarrow{D} N(0,1)$ but does $\sqrt{\frac{J(\hat{\theta})}{J(\theta)}} \xrightarrow{P} 1$ (then we could apply (D2)/Slutsky's).
   Note that
$$\sqrt{\frac{J(\hat{\theta})}{J(\theta)}} = \sqrt{\frac{\frac{n}{\hat{\theta}^2}}{\frac{n}{\theta^2}}} = \frac{\theta}{\hat{\theta}} \xrightarrow{D} 1$$

as $\hat{\theta} \xrightarrow{P} \theta$ (by (P1) we have $g(x) = \frac{\theta}{x}$ which is continuous at θ then $g(\hat{\theta}) = \frac{\theta}{\hat{\theta}} \xrightarrow{P} g(\theta) = \frac{\theta}{\theta} = 1$).

4. Verify $\sqrt{I(\hat{\theta})}(\hat{\theta} - \theta) \xrightarrow{D} N(0,1)$.
   Recall $I(\hat{\theta}) = J(\hat{\theta}) = \frac{n}{\hat{\theta}^2}$.
   Follow the same steps as (3).

Exercise 29.1. Verify the four asymptotic properties for $X_i \overset{iid}{\sim} POI(\mu)$ for $i = 1, \ldots, n$.

Remark 29.1. From the asymptotic property (2) where $\sqrt{J(\theta)}(\hat{\theta} - \theta) \xrightarrow{D} Z \sim N(0,1)$ we see that for large n we have approximately

$$\hat{\theta} - \theta \sim N(0, \frac{1}{J(\theta)})$$

That is $Var(\hat{\theta}) \approx \frac{1}{J(\theta)}$. The larger value of $J(\theta)$ (i.e. more information) the smaller $Var(\hat{\theta})$ i.e. $\hat{\theta}$ is a more precise estimator

30 November 21, 2018

30.1 Another example of verifying asymptotic properties

Example 30.1. Suppose $X_i \overset{iid}{\sim} WEI(\theta, 2)$, $\theta > 0$. Note that

$$E[x_i^k] = \theta^k \Gamma(\frac{k}{2} + 1) \qquad k = 1, 2, \ldots$$

Verify the four asymptotic properties.

Solution. We skip the calculations of $L(\theta)$ and $l(\theta)$ and jump to

$$S(\theta) = \frac{-2n}{\theta} + \frac{2\sum_{i=1}^n X_i^2}{\theta^3}$$

solving $S(\theta) = 0$ we get $\hat{\theta} = \sqrt{\frac{\sum_{i=1}^n X_i^2}{n}}$ for $\Omega = \mathbb{R}^+$.

Exercise 30.1. Verify MLE with 1st derivative test.

1. Verify $\hat{\theta} \xrightarrow{P} \theta$.

   Let $Y_i = X_i^2$ therefore $\hat{\theta} = \sqrt{\frac{\sum_{i=1}^n Y_i}{n}} = \sqrt{\bar{Y}}$.

   Since $X_i \overset{iid}{\sim} WEI(\theta, 2)$ then Y_i's are IID and

$$\begin{aligned} E(Y_i) &= E(X_i^2) = \theta^2 \Gamma(\frac{2}{2} + 1) = \theta^2 \Gamma(2) = \theta^2 < \infty \\ E(Y_i^2) &= E(X_i^4) = \theta^4 \Gamma(3) = 2\theta^4 \\ Var(Y_i) &= E(Y_i^2) - E(Y_i)^2 = 2\theta^4 - \theta^4 = \theta^4 < \infty \end{aligned}$$

   Due to WLLN, we have $\bar{Y} \xrightarrow{P} E(Y_i) = \theta^2$.

   Let $g(x) = \sqrt{x}$ which is continuous for all $x > 0$ so it is continuous at θ^2.

Therefore according to (P1) we have

$$g(\bar{Y}) = \sqrt{\bar{Y}} = \hat{\theta} \xrightarrow{P} g(\theta^2) = \theta$$

2. Verify $\sqrt{J(\theta)}(\hat{\theta} - \theta) \xrightarrow{D} N(0, 1)$.

   Note $S(\theta) = \frac{-2n}{\theta} + \frac{2\sum X_i^2}{\theta^3}$ and $I(\theta) = \frac{6\sum X_i^2}{\theta^4} - \frac{2n}{\theta^2}$.

   So we have

$$\begin{aligned} J(\theta) &= E(I(\theta; X)) \\ &= E\big(\frac{6\sum X_i^2}{\theta^4} - \frac{2n}{\theta^2}\big) \\ &= \frac{6\sum E(X_i^2)}{\theta^4} - \frac{2n}{\theta^2} \\ &= \frac{6\sum n\theta^2}{\theta^4} - \frac{2n}{\theta^2} \\ &= \frac{4n}{\theta^2} \end{aligned}$$

   Therefore $\sqrt{J(\theta)}(\hat{\theta} - \theta) = \sqrt{\frac{4n}{\theta^2}}(\sqrt{\bar{Y}} - \sqrt{\theta^2})$.

   Due to the CLT we have

$$\begin{aligned} &\frac{\sqrt{n}(\bar{Y} - E(Y_i))}{\sqrt{Var(Y_i)}} \xrightarrow{D} N(0, 1) \\ \Rightarrow &\frac{\sqrt{n}(\bar{Y} - \theta^2)}{\theta^2} \xrightarrow{D} N(0, 1) \end{aligned}$$

   so

$$\theta^2 \frac{\sqrt{n}(\bar{Y} - \theta^2)}{\theta^2} \xrightarrow{D} \theta^2 Z \sim N(0, \theta^4)$$

   due to (D1): recall if $Z_n \xrightarrow{D} Z$ then $g(Z_n) \xrightarrow{D} g(Z)$ if $g(\cdot)$ is continuous on the entire support set of Z. Let $g(x) = \theta^2 x$ which is continuous for all $x \in \mathbb{R}$ so it is continuous on the entire support set of Z.

   Therefore $\sqrt{n}(\bar{Y} - \theta^2) \xrightarrow{D} N(0, \theta^4)$.

   Let $g(x) = \sqrt{x}$ where $g'(x) = \frac{1}{2}x^{-\frac{1}{2}}$ for all $x > 0$ therefore $g'(\theta^2) = \frac{1}{2\theta} \neq 0$ for all $\theta > 0$.

   Due to the Delta method we have

$$\begin{aligned} &\sqrt{n}\big(g(\bar{Y}) - g(\theta^2)\big) \xrightarrow{D} N(0, (g'(\theta^2))^2\theta^4) \\ \Rightarrow &\sqrt{n}\big(\sqrt{\bar{Y}} - \sqrt{\theta^2}\big) \xrightarrow{D} N(0, \frac{\theta^2}{4}) \end{aligned}$$

Then due to (D1) we have

$$
\begin{aligned}
&(\frac{2}{\theta})\sqrt{n}(\sqrt{\bar{Y}} - \sqrt{\theta^2}) \xrightarrow{D} (\frac{2}{\theta})N(0, \frac{\theta^2}{4}) \\
\Rightarrow&\sqrt{\frac{4n}{\theta^2}}(\sqrt{\bar{Y}} - \sqrt{\theta^2}) \xrightarrow{D} N(0,1)
\end{aligned}
$$

as desired.

3. Exercise.
4. Exercise.

30.2 Set estimation

We attempt to find a set of possible values of θ. For θ being a scalar, we exclusively focus on **confidence intervals** (most commonly used type of set estimation).

Remark 30.1. Likelihood intervals is also another type of set estimation, but will not be covered.

Question 30.1. How do we construct a confidence interval (CI)? We use a **Pivotal Quantity** or **asymptotic Pivotal Quantity**.

Example 30.2. Suppose $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$ $\mu \in \mathbb{R}$ and $\sigma > 0$.
We use

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$$

which we proved in an earlier theorem to construct a CI for μ. This is our pivotal quantity.
We then find the set of μ's such that

$$
\begin{aligned}
&P\big(t_{0.05}(n-1) \leq \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \leq t_{0.95}(n-1)\big) = 0.90 \\
\Rightarrow&P\big(A(X) = \bar{X} + t_{0.05}(n-1)\frac{S}{\sqrt{n}} \leq \mu \leq B(X) = \bar{X} - t_{0.95}(n-1)\frac{S}{\sqrt{n}}\big) = 0.90
\end{aligned}
$$

Then $[A(X), B(X)]$ is an equal-tail 90% CI for μ.

Definition 30.1 (Pivotal Quantity)**.** A quantity $Q(X;\theta)$ is called a **Pivotal Quantity** if the distribution of $Q(X;\theta)$ *does NOT* depend on any unknown parameter(s).

Example 30.3. Suppose $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $\mu \in \mathbb{R}, \sigma > 0$.

$Q(X;\theta)$

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Pivotal Quantity? Yes since $N(0,1)$ is non-parameterized.

CI? For μ we cannot create a CI since σ is unknown.

$Q(X;\theta)$

$$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$$

Pivotal Quantity? Yes since $t(n-1)$ is non-parameterized.

CI? Yes for μ.

$Q(X;\theta)$

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

Pivotal Quantity? Yes since $\chi^2(n-1)$ is non-parameterized.

CI? Yes for σ^2.

$Q(X;\theta)$

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$$

Pivotal Quantity? Yes since $\chi^2(n)$ is non-parameterized.

CI? No for σ^2 since μ is unknown.

Definition 30.2 (Asymptotic Pivotal Quanity)**.** A quantity $Q(X;\theta)$ is called an **asymptotic Pivotal Quantity** if the limiting distribution of $Q(X;\theta)$ *does NOT* depend on any unknown parameter(s).

Remark 30.2. Note $\sqrt{J(\theta)}(\hat{\theta}-\theta) \xrightarrow{D} N(0,1)$ is an example of an asymptotic PQ.

31 November 23, 2018

31.1 Applying asymptotic properties to confidence intervals

Example 31.1. Suppose $X_i \overset{iid}{\sim} POI(\theta)$. Find an approximate equal-tail CI for θ.
Recall

$$\sqrt{J(\theta)}(\hat{\theta}-\theta) \xrightarrow{D} N(0,1)$$
$$\sqrt{J(\hat{\theta})}(\hat{\theta}-\theta) \xrightarrow{D} N(0,1)$$
$$\sqrt{I(\hat{\theta})}(\hat{\theta}-\theta) \xrightarrow{D} N(0,1)$$

which are all (except the first one) free of any unknown parameter(s): they are all asymptotic pivotal quantities $Q(x;\theta)$.
For example

$$P(z_{0.05} \le \sqrt{J(\hat{\theta})}(\hat{\theta}-\theta) \le z_{0.95}) \approx 0.90$$
$$\Rightarrow P(\hat{\theta} + z_{0.05}\frac{1}{\sqrt{J(\hat{\theta})}} \le \theta \le \hat{\theta} + z_{0.95}\frac{1}{\sqrt{J(\hat{\theta})}}) \approx 0.90$$

where $\hat{\theta} = \bar{x}$, $z_{0.95} = 1.645$, $z_{0.05} = -1.645$ and $J(\hat{\theta}) = \frac{n}{\hat{\theta}}$.
Notice that $J(\theta) = \frac{n}{\theta}$, $I(\theta) = \frac{n}{\theta}$ and $I(\hat{\theta}) = J(\hat{\theta}) = \frac{n}{\hat{\theta}}$.
However, we cannot use $\sqrt{J(\theta)}(\hat{\theta} - \theta) \xrightarrow{D} N(0,1)$ since $J(\theta)$ contains an unknown parameter θ.

31.2 Pivotal Quantities in general

How do we find pivotal quantities in general?

Theorem 31.1. Suppose $X_i \overset{iid}{\sim} f(x;\theta)$, $i = 1, \ldots, n$. Let $\hat{\theta}$ be the ML estimator of θ.

1. If θ is a location parameter, then $(\hat{\theta} - \theta)$ is a PQ.
2. If θ is a scale parameter, then $\frac{\hat{\theta}}{\theta}$ is a PQ.

Remark 31.1. If θ is neither a location nor a scale parameter, we can use the asymptotic properties based on the MLE properties to derive a CI for θ.

Example 31.2. Suppose $X_i \overset{iid}{\sim} N(\mu, 1)$. Find an equal-tail 90% CI for μ.

Solution. Verify μ is a location parameter i.e. $f(x;\mu) = f(x - \mu; 0)$.
Note that

$$\begin{aligned}
f(x;\mu) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} \\
f(x;0) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \\
\Rightarrow f(x-\mu;0) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} = f(x;\mu)
\end{aligned}$$

so μ is a location parameter.
Note that $\hat{\mu} = \bar{X}$ and so $\hat{\mu} - \mu$ is a PQ.
We have $\hat{\mu} - \mu = \bar{X} - \mu \sim N(0, \frac{1}{n})$ (which is free of any unknown parameters) since $X_i \overset{iid}{\sim} N(\mu, 1)$ and $\bar{X} \sim N(\mu, \frac{1}{n})$.
Therefore $\sqrt{n}(\bar{X} - \mu) \sim N(0,1)$ then

$$\begin{aligned}
&P(z_{0.05} \le \sqrt{n}(\bar{X} - \mu) \le z_{0.95}) = 0.90 \\
&\Rightarrow P(\bar{X} - 1.645\frac{1}{\sqrt{n}} \le \mu \le \bar{X} + 1.645\frac{1}{\sqrt{n}}) = 0.90
\end{aligned}$$

Example 31.3. Suppose $X_i \overset{iid}{\sim} N(0, \sigma^2)$, $\sigma > 0$. Find an equal-tail 90% CI for σ.

Solution. Verify σ is a scale parameter i.e. $f(x;\sigma) = \frac{1}{\sigma} f(\frac{x}{\sigma}; 1)$?

$$\begin{aligned}
f(x;\sigma) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \\
f(x;1) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \\
\Rightarrow \frac{1}{\sigma} f(\frac{x}{\sigma};1) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\frac{x}{\sigma}\right)^2}{2}} = f(x;\sigma)
\end{aligned}$$

therefore σ is a scale parameter and $\frac{\hat{\sigma}}{\sigma}$ is a PQ.

Exercise 31.1. Show $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n X_i^2}{n}}$.
Therefore

$$\frac{\hat{\sigma}}{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^n \big(\frac{X_i}{\sigma}\big)^2}$$

is a PQ.
Note that $\sum_{i=1}^n \big(\frac{X_i}{\sigma}\big)^2 \sim \chi^2(n)$ for $X_i \overset{iid}{\sim} N(0,\sigma^2)$ (recall if $X_i \overset{iid}{\sim} N(\mu,\sigma^2)$ then $\sum_{i=1}^n \big(\frac{X_i-\mu}{\sigma}\big)^2 \sim \chi^2(n)$) then we use

$$n\big(\frac{\hat{\sigma}}{\sigma}\big)^2 = \sum_{i=1}^n \big(\frac{X_i}{\sigma}\big)^2 \sim \chi^2(n)$$

Therefore

$$\begin{aligned} &P(\chi^2(n)_{0.05} \le n\big(\frac{\hat{\sigma}}{\sigma}\big)^2 \le \chi^2(n)_{0.95}) = 0.90 \\ &\Rightarrow P(A(x) = \frac{n\hat{\sigma}^2}{\chi^2(n)_{0.05}} \le \sigma^2 \le B(x) = \frac{n\hat{\sigma}^2}{\chi^2(n)_{0.95}}) = 0.90 \end{aligned}$$

where $[\sqrt{A(x)}, \sqrt{B(x)}]$ is an equal-tail 90% CI for σ.

Example 31.4. Suppose $X_i \overset{iid}{\sim} EXP(1,\theta)$. Find an equal-tail 90% CI for θ.

Solution. Verify θ is a location parameter.

$$\begin{aligned} f(x;\theta) &= e^{-(x-\theta)} \qquad x \ge \theta \\ L(\theta) &= \prod_{i=1}^n f(x_i;\theta) = e^{-\sum_{i=1}^n x_i} e^{n\theta} \qquad x_i \ge \theta \quad \forall i = 1,\dots,n \end{aligned}$$

To maximize $L(\theta)$ we take $\hat{\theta} = \text{argmax}_{\theta\in\Omega} L(\theta)$ i.e. we want to maximize $e^{n\theta}$ i.e. we require θ to be as large as possible.
Since $x_i \ge \theta$ for all $i = 1,\dots,n$ $\hat{\theta} = \min(x_1,\dots,x_n)$.
Therefore $\hat{\theta} - \theta$ is a PQ, where

$$P(q_{0.05} \le \hat{\theta} - \theta \le q_{0.95}) = 0.90$$

where $q_{0.05}$ and $q_{0.95}$ are the 5% and 95% quantile of the distribution of $\hat{\theta} - \theta$.
Let $Q(q) = P(\hat{\theta} - \theta \le q)$, cdf of $\hat{\theta} - \theta$. Note that

$$\begin{aligned} Q(q) &= P(\min(X_1,\dots,X_n) \le q + \theta) \\ &= 1 - P(\min(X_1,\dots,X_n) > q + \theta) \\ &= 1 - P(X_1 > q + \theta)\dots P(X_n > q + \theta) \\ &= 1 - \big(P(X_1 > q + \theta)\big)^n \end{aligned}$$

Recall for $X_i \overset{iid}{\sim} EXP(1,\theta)$ we have

$$F(x) = P(X_i \le x) = \begin{cases} 0 & \text{if } x < \theta \\ 1 - e^{-(x-\theta)} & \text{if } x \ge \theta \end{cases}$$

so we have

$$Q(q) = 1 - \big(1 - F(\theta + q)\big)^n$$
$$= 1 - (e^{-(\theta+q-\theta)})^n$$
$$= 1 - e^{-nq}$$

Note we require $\theta + q \geq \theta$ for $F(\cdot)$ to be non-zero. Recall we initially had $\hat{\theta} = \min(X_1, \ldots, X_n) \geq \theta$ so $\hat{\theta} - \theta \geq 0$ so

$$Q(q) = \begin{cases} 0 & \text{if } q < 0 \\ 1 - e^{-nq} & \text{if } q > 0 \end{cases}$$

32 November 26, 2018

32.1 Un-equal tail CIs example

Suppose $X_i \overset{iid}{\sim} f(x;\theta) = e^{-(x-\theta)}$, $x \geq 0$. Recall $\hat{\theta} = \min(x_1, \ldots, x_n) \geq \theta$. Find a 90% CI for θ.

Solution. Note that $P(q_{0.05} \leq \hat{\theta} - \theta \leq q_{0.95}) = 0.90$.
Recall for X being a *continuous* r.v. with cdf $F(\cdot)$ we have $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$.
So we have $P(\hat{\theta} - \theta \leq q_{0.95}) - P(\hat{\theta} - \theta \leq q_{0.05}) = 0.90$ or $Q(q_{0.95}) - Q(q_{0.05}) = 0.90$.
Note that $Q(q_{0.95}) = 1 - e^{-nq_{0.95}}$ where $Q(q_{0.95}) = 0.95$ (we choose this so that we have an equal-tail) so

$$1 - e^{-nq_{0.95}} = 0.95 \Rightarrow q_{0.95} = \frac{1}{n} \log 0.05$$

and also $q_{0.05} = -\frac{1}{n} \log 0.95$.
Therefore $P(-\frac{1}{n} \log 0.95 \leq \hat{\theta} - \theta \leq \frac{1}{n} \log 0.05) = 0.90$ that is $P(A(x) = \hat{\theta} + \frac{1}{n} \log 0.05 \leq \theta \leq B(x) = \hat{\theta} + \frac{1}{n} \log 0.95) = 0.90$. such that $[A(x), B(x)]$ is an equal-tail 90% CI for θ.
Notice that $\frac{1}{n} \log 0.05, \frac{1}{n} \log 0.95 < 0$ therefore $\hat{\theta}$ is outside of $[A(x), B(x)]$, which is **unfortunate**.
Recall $\hat{\theta}$ is the single most possible value of θ where $\hat{\theta} = \text{argmax}_{\theta \in \Omega} L(\theta)$.
Since $\hat{\theta} \geq \theta$ then $\hat{\theta} - \theta \geq 0$. So instead of $P(q_{0.05} \leq \hat{\theta} - \theta \leq q_{0.95}) = 0.90$ we can use

$$P(0 \leq \hat{\theta} - \theta \leq q_{0.90}) = 0.90$$

i.e. a non-equal tail CI, so we have

$$P(0 \leq \hat{\theta} - \theta \leq -\frac{1}{n} \log 0.1) = 0.90$$
$$\Rightarrow P(A(x) = \hat{\theta} + \frac{1}{n} \log 0.1 \leq \theta \leq B(x) = \hat{\theta}) = 0.9$$

such that $[A(x), B(x)]$ is a 90% CI for θ that contains $\hat{\theta}$ and is a *better* CI for θ than the equal-tail one.

32.2 Point estimation vs set estimation

Example 32.1. Suppose $X_i \overset{iid}{\sim} N(\mu, 1)$, $i = 1, \ldots, n$. where $\bar{X}$ has a continuous distribution i.e. $P(\bar{X} = \mu) = 0$. Along with a point estimate of any unknown parameter we want to understand how close $\hat{\mu}$ is to the true value of the unknown parameter μ by providing **MSE: mean squared error** where

$$MSE = E\big[(\hat{\theta} - \theta)^2\big]$$

Another approach is set esetimation: we can consider the *probability* that such a set will contain the true value of the unknown parameter μ, i.e. a measure of "accuracy".
That is point estimation is precision and set estimation is accuracy.

32.3 Multiparameter MLE

Given $\vec{\theta} = (\theta_1, \ldots, \theta_k)^T$ we have $L(\vec{\theta}) = L(\theta_1, \ldots, \theta_k)$ and $l(\vec{\theta}) = l(\theta_1, \ldots, \theta_k)$. Let $S(\vec{\theta}) = \left(\frac{\partial l(\vec{\theta})}{\partial \theta_1}, \ldots, \frac{\partial l(\vec{\theta})}{\partial \theta_k}\right)^T$.
To find $\hat{\vec{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)^T$, we solve for $S(\vec{0}) = (0, \ldots, 0)^T$ that is

$$\begin{aligned}\frac{\partial l(\vec{\theta})}{\partial \theta_1} &= 0 \\ &\vdots \\ \frac{\partial l(\vec{\theta})}{\partial \theta_k} &= 0\end{aligned}$$

simultaneously.
We then must carry out a multivariate 2nd deriviate test to verify $\hat{\vec{\theta}}$ is indeed a maximizer.
Recall in the 1-D case if θ is a scalar the univariate 2nd derivative test verifies $\left.\frac{\partial^2 l(\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}} < 0$, or if $I(\theta) = -\frac{\partial^2 l(\theta)}{\partial \theta^2}$ we verify $I(\theta) > 0$.
In the multivariate 2nd derivative test we verify that $I(\hat{\theta})$ is **positive definite** that is for

$$I(\theta) = \begin{bmatrix} -\frac{\partial^2 l(\vec{\theta})}{\partial \theta_1^2} & \cdots & -\frac{\partial^2 l(\vec{\theta})}{\partial \theta_1 \partial \theta_k} \\ \vdots & \ddots & \vdots \\ -\frac{\partial^2 l(\vec{\theta})}{\partial \theta_k \partial \theta_1} & \cdots & -\frac{\partial^2 l(\vec{\theta})}{\partial \theta_k^2} \end{bmatrix}$$

where $I \in \mathbb{R}^{k \times k}$. Note that $I(\theta)$ is a symmetric matrix where

$$-\frac{\partial^2 l(\vec{\theta})}{\partial \theta_i \partial \theta_j} = -\frac{\partial^2 l(\vec{\theta})}{\partial \theta_j \partial \theta_i}$$

Remark 32.1. A matrix A is positive definite if $\vec{x}^T A \vec{x} > 0$ for $\vec{x} \neq \vec{0}$.

33 November 28, 2018

33.1 MLE for both mean and variance in normal distribution

Example 33.1. Suppose $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$.
Find the MLE of (μ, σ).

Solution. Let $\vec{\theta} = (\mu, \sigma)^T$. Note that

$$\begin{aligned} L(\mu, \sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} \\ &= (2\pi)^{-\frac{n}{2}} (\sigma)^{-n} e^{-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2}} \end{aligned}$$

So

$$l(\mu,\sigma) = -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

Note that

$$S(\mu,\sigma) = \begin{bmatrix} \frac{\partial l(\mu,\sigma)}{\partial \mu} \\ \frac{\partial l(\mu,\sigma)}{\partial \sigma} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{n}{\sigma} \end{bmatrix}$$

now we solve for $S(\mu,\sigma) = (0,0)^T$ simultaneously to find $(\hat{\mu},\hat{\sigma})$ that is

$$\frac{1}{\sigma^2}\Big(\sum_{i=1}^{n} X_i - n\mu\Big) = 0 \Rightarrow \hat{\mu}\bar{X}$$
$$\frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{n}{\sigma} = 0 \Rightarrow \frac{1}{\sigma^2}\sum(X_i-\mu)^2 = n$$

so $\hat{\sigma} = \sqrt{\frac{\sum(X_i-\bar{X})^2}{n}}$.
Next we verify $I(\hat{\mu},\hat{\sigma})$ is positive definite.
Note that

$$I(\mu,\sigma) = \begin{bmatrix} -\frac{\partial^2 l(\mu,\sigma)}{\partial \mu^2} & -\frac{\partial^2 l(\mu,\sigma)}{\partial \mu \partial \sigma} \\ -\frac{\partial^2 l(\mu,\sigma)}{\partial \sigma \partial \mu} & -\frac{\partial^2 l(\mu,\sigma)}{\partial \sigma^2} \end{bmatrix}$$

we have

$$-\frac{\partial^2 l(\mu,\sigma)}{\partial \mu^2} = \frac{\partial}{\partial \mu}\Big(\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)\Big) = \frac{n}{\sigma^2}$$
$$-\frac{\partial^2 l(\mu,\sigma)}{\partial \mu \partial \sigma} = \frac{\partial}{\partial \sigma}\Big(\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)\Big) = \frac{2}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)$$
$$-\frac{\partial^2 l(\mu,\sigma)}{\partial \sigma^2} = \frac{\partial}{\partial \sigma}\Big(-\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2\Big) = -\frac{n}{\sigma^2} + \frac{3\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^4}$$

Therefore

$$I(\mu,\sigma) = \begin{bmatrix} \frac{n}{\sigma^2} & \frac{2}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu) \\ \frac{2}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu) & -\frac{n}{\sigma^2} + \frac{3\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^4} \end{bmatrix}$$

so

$$I(\hat{\mu},\hat{\sigma}) = \begin{bmatrix} \frac{n}{\hat{\sigma}^2} & \frac{2}{\hat{\sigma}^3}\sum_{i=1}^{n}(X_i-\hat{\mu}) \\ \frac{2}{\hat{\sigma}^3}\sum_{i=1}^{n}(X_i-\hat{\mu}) & -\frac{n}{\hat{\sigma}^2} + \frac{3\sum_{i=1}^{n}(X_i-\hat{\mu})^2}{\hat{\sigma}^4} \end{bmatrix}$$

note that

$$\frac{2}{\hat{\sigma}^3}\sum_{i=1}^{n}(X_i-\hat{\mu}) = \frac{2}{\hat{\sigma}^3}\Big(\sum_{i=1}^{n} X_i - n\bar{X}\Big) = 0$$

also for $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}}$ we have

$$-\frac{n}{\hat{\sigma}^2} + \frac{3\sum_{i=1}^{n}(X_i-\hat{\mu})^2}{\hat{\sigma}^4} = \frac{2n}{(\hat{\sigma})^2}$$

so indeed

$$I(\hat{\mu}, \hat{\sigma}) = \begin{bmatrix} \frac{n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{2n}{(\hat{\sigma})^2} \end{bmatrix}$$

which is obviously positive definite since diagonal entries are strictly positive therefore eigenvalues are positive so $(\hat{\mu}, \hat{\sigma})$ is the MLE of (μ, σ).

33.2 Asymptotic properties of multiparameter MLE

1. $\vec{\hat{\theta}} \xrightarrow{P} \vec{\theta}$
2. $\left(J(\vec{\theta})\right)^{\frac{1}{2}}(\vec{\hat{\theta}} - \vec{\theta}) \xrightarrow{D} \vec{Z} \sim MVN((0, \ldots, 0)^T, I_{k\times k})$
3. $\left(J(\vec{\hat{\theta}})\right)^{\frac{1}{2}}(\vec{\hat{\theta}} - \vec{\theta}) \xrightarrow{D} \vec{Z} \sim MVN(0_k, I_{k\times k})$
4. $\left(I(\vec{\hat{\theta}})\right)^{\frac{1}{2}}(\vec{\hat{\theta}} - \vec{\theta}) \xrightarrow{D} \vec{Z} \sim MVN(0_k, I_{k\times k})$

Remark 33.1. If A is a positive definite matrix then $A = U\Lambda U^T$ where $U = (\vec{u_1}, \ldots, \vec{u_k})_{k\times k}$ and

$$\Gamma = \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_k \end{bmatrix}$$

such that $Au_i = \lambda_i u_i$ for all $i = 1, \ldots, k$ (i.e. λ_i and u_i are the ith eigenvalue and eigenvectors of A).
Then $A^{\frac{1}{2}} = U\Gamma^{\frac{1}{2}}U^T$ where

$$\Gamma^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sqrt{\lambda_k} \end{bmatrix}$$

In this course we are not responsible for verifying the four properties for multiparameter MLEs.

33.3 Newton's method for multiparameters

Newton's method for multiple parameters $\vec{\theta}$ is

$$\vec{\theta}^{(r+1)} = \vec{\theta}^{(r)} + \left(I(\vec{\theta}^{(r)})\right)^{-1} S(\vec{\theta}^{(r)})$$

33.4 Hypothesis testing

Definition 33.1 (Null hypothesis)**.** The **null hypothesis** $H_0 : \theta = \theta_0$ i.e. the unknown parameter is equal to the pre-specified θ_0.

Example 33.2.

1. Is the average of STAT 330 midterm 3 80%? We let $H_0 : \mu = 80\%$.
2. Did STAT330 section 001 and 002 perform equally as well in midterm 3? We let $H_0 : \mu_1 - \mu_2 = 0$.

Remark 33.2. The null hypothesis generally corresponds the the "yes" part of the question and respresents prior knowledge i.e. status quo.

Definition 33.2 (Test statistic)**.** A **test statistic** $T(X)$ to measure the evidence supporting/against H_0. $T(X)$ should typically have these two desired properties:

1. The larger the value of T implies more evidence against H_0
2. The smaller the value of T implies more evidence supporting H_0

Example 33.3. Suppose $X_i \overset{iid}{\sim} N(\mu, 1)$. To test $H_0 : \mu = 0$ we use $T(X) = \frac{|\bar{X}|}{\frac{1}{\sqrt{n}}}$.
Notice that for a given sample size n, $T(X)$ is large if $\bar{X}$ is large. If $\mu = 0$ then $\bar{X}$ should be small, otherwise $\bar{X}$ will be large.
Therefore if $T(X)$ is large it indicates $\mu \neq 0$ (and if it's small it indicates $\mu = 0$).

Definition 33.3 (p-value)**.** The p-value is defined as $P_{H_0}(T(X) \geq T(x))$ i.e. the probability of $T(X)$ being at least $T(x)$ under H_0, where $T(x)$ is the observed value of $T(X)$ given the observed sample x.
Recall: if $T(x)$ is large we have evidence against H_0 therefore $P_{H_0}(T(X) \geq T(x))$ is small i.e. a small p-value leads to evidence against H_0.

34 November 30, 2018

Definition 34.1 (Significance level)**.** The **significance level** α is such that we reject H_0 if the p-value $< \alpha$ (i.e. p-value is small enough i.e. we have evidence strong enough against H_0).

Remark 34.1. In practice we choose $\alpha = 0.05$ as a convention.

34.1 Likelihood ratio test for simple null hypothesis

Definition 34.2 (Simple hypothesis)**.** A hypothesis is called a **simple hypothesis** if it completely specifies the distribution of the random sample $(X_1, \ldots, X_n)$ where $X_i \overset{iid}{\sim} f(x; \theta)$.
i.e. the distribution of $(X_1, \ldots, X_n)$ is free of any unknown parameter(s).

Example 34.1. Suppose $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$.

1. Let $H_0 : \mu = 0$ and $\sigma = 1$ (where $H_a : \mu \neq 0$ or $\sigma \neq 1$).

   Note that H_0 is a simple null hypothesis (since it specifies all unknown parmemeters whereas H_a is a composite alternative hypotehsis since it does not specify all unknown parmeters).

2. Let $H_0 : \mu = 0$ (and $H_a : \mu \neq 0$).

   Both H_0 and H_a are composite hypotheses.

3. Let $H_0 : \sigma = 1$ ($H_a : \sigma \neq 1$).

   Both are also composite hypotheses.

   Both H_0 and H_a are composite hypotheses.

Definition 34.3 (Likelihood ratio test statistic)**.** A **likelihood ratio test statistic for** $H_0 : \theta = \theta_0$ **(simple hypothesis) vs.** $H_a : \theta \neq \theta_0$ **is defined as**

$$\begin{aligned}\lambda(X) &= 2 \log \frac{L(\hat{\theta}; X)}{L(\theta_0; X)} \\ &= 2(l(\hat{\theta}; X) - l(\theta_0; X))\end{aligned}$$

Remark 34.2. 1. $\frac{L(\hat{\theta}; X)}{L(\theta_0; X)} \geq 1$ as $\hat{\theta}$ is the MLE.

2. From (1) $\lambda(X) \geq 0$

3. Under some regularity conditions, $\hat{\theta} \xrightarrow{P} \theta$ under $H_0 : \theta = \theta_0$ i.e. $\hat{\theta} \xrightarrow{P} \theta_0$ under H_0, so $\frac{L(\hat{\theta};X)}{L(\theta_0;X)} \to 1$ from above for large n under H_0 i.e. $\lambda(X)$ should be close to 0 under H_0.

4. Under some regularity conditions (STAT 450) we can show that $\lambda(X) \xrightarrow{D} W \sim \chi^2(k)$ under H_0 where $\theta = (\theta_1, \dots, \theta_k)^T$.

   Therefore the p-value $= P_{H_0}(\lambda(X) \geq \lambda(x)) \approx P(W \geq \lambda(x))$.

Example 34.2. Suppose $X_i \overset{iid}{\sim} N(\mu, 1)$, $\mu \in \mathbb{R}$.
We test $H_0 : \mu = 0$ vs $H_a = \mu \neq 0$ for $n = 25$ and $\bar{x} = \frac{1}{2}$ using the LRT statistic.

Solution.

$$\begin{aligned}\lambda(X) &= 2 \log \frac{L(\hat{\mu}; X)}{L(\mu = 0; X)} \\ &= 2(l(\hat{\mu}; X) - l(\mu = 0; X))\end{aligned}$$

Note that

$$\begin{aligned}L(\mu; X) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(X_i-\mu)^2}{2}} = (2\pi)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2}} \\ \Rightarrow l(\mu; X) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)^2\end{aligned}$$

Exercise 34.1. Show $\hat{\mu} = \bar{X}$.

Thus

$$\begin{aligned}\lambda(X) &= 2\big(-\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(X_i - \bar{X})^2 - \big(-\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(X_i - 0)^2\big)\big) \\ &= \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= n(\bar{X})^2\end{aligned}$$

i.e. $\lambda(x) = n(\bar{x})^2$. Recall the p-value is $P(W \geq \lambda(x))$ as $W \sim \chi^2(1)$ and $\lambda(X) \xrightarrow{D} W \sim \chi^2(1)$ under H_0.
So our p-value is $P(W \geq (25)(1/2)^2) = P(W \geq 6.25)$.
We reject H_0 if p-value $< \alpha = 0.05$.
Note that $P(W \geq 3.84) = 0.05$ for $W \sim \chi^2(1)$, therefore $P(W \geq 6.25) < P(W \geq 3.84) = 0.05$ so we reject H_0.

Remark 34.3. In the past for $H_0 : \mu = 0$ for $X_i \overset{iid}{\sim} N(\mu, 1)$ we used the statistic

$$T(X) = \frac{|\bar{X}|}{\frac{1}{\sqrt{n}}}$$

and rejected H_0 if $T(X)$ is too large (p-value is too small).
Recall $\lambda(X) = n(\bar{X})^2 = \frac{\bar{X}^2}{\frac{1}{n}} = \big(\frac{|\bar{X}|}{\frac{1}{\sqrt{n}}}\big)^2 = (T(X))^2$.
for LRT statistic we also reject H_0 if $\lambda(x)$ is too large i.e. iff $T(x) = \frac{|\bar{x}|}{\frac{1}{\sqrt{n}}}$ is too large.

Exercise 34.2. Suppose $X_i \overset{iid}{\sim} N(0, \sigma^2)$. We test $H_0 : \sigma = 1$ for $n = 25$ and $\sum_{i=1}^n x_i^2 = 30$.

34.2 LRT with multinomial distributions

Example 34.3. Suppose $(X_1, X_2) \sim MULT(n, \theta^2, 2\theta(1-\theta))$ where $0 < \theta < 1$.
We test $H_0 : \theta = \frac{1}{2}$.

Solution.

$$L(\theta; X) = \frac{n!}{X_1!X_2!(n - X_1 - X_2)!}(\theta^2)^{X_1}(2\theta(1-\theta))^{X_2}((1-\theta)^2)^{n-X_1-X_2}$$

where $\theta^2 + 2\theta(1-\theta) + (1-\theta)^2 = 1$ (i.e. $p_1 + p_2 + (1 - p_1 - p_2) = 1$).
We also have

$$l(\theta; X) = \log\big(\frac{n!}{X_1!X_2!(n - X_1 - X_2)!}2^{X_2}\big) + (2X_1 + X_2)\log\theta + (2n - 2X_1 - 2X_2)\log(1-\theta)$$

Exercise 34.3. Show $\hat{\theta} = \frac{(2X_1 + X_2)}{2n}$.

So we have

$$\lambda(X) = 2\big((2X_1 + X_2)\log(\hat{\theta}) + (2n - 2X_1 - X_2)\log(1-\hat{\theta}) - 2n\log(1/2)\big)$$

we can substitute in $n = 20, X_1 = 5$ and $X_2 = 6$ (values from our samples) which yields $\lambda(x) = 1.61$.
Thus the p-value is $\approx P(W \geq 1.61) > P(W \geq 3.84) = 0.05$.
Therefore we do not reject $H_0 : \sigma = \frac{1}{2}$.

35 December 3, 2018

35.1 LRT for a composite null hypothesis

Given that $X_i \overset{iid}{\sim} f(x, \theta)$, $\theta \in \Omega$ and $f(x; \theta)$ satisfies regularity conditions:

Definition 35.1 (LRT for multiparemeter)**.** Suppose we are testing $H_0 : \theta \in \Omega_0$, $\theta \in \mathbb{R}^{k\times 1}$ (k-dimensional vector) vs $H_a : \theta \in \Omega \setminus \Omega_0$ where Ω_0 is the parameter space under H_0.
The **LRT statistic** (for multi-dimensional pararameter θ) is defined as

$$\lambda(X) = 2\log\frac{L(\hat{\theta}; X)}{L(\tilde{\theta}; X)}$$

where $\hat{\theta} = \text{argmax}_\Omega L(\theta; X)$ (max θ over the entire parameter space Ω) and $\tilde{\theta} = \text{argmax}_{\Omega_0} L(\theta; X)$ (max θ over just our null hypothesis parameter space Ω_0).

Remark 35.1. $\hat{\theta}$ is an **unconstrained MLE** of θ whereas $\tilde{\theta}$ is a **constrained MLE** of θ.

Since Ω_0 is a subset of Ω then $L(\hat{\theta}; X) \geq L(\tilde{\theta}; X)$ i.e. $\lambda(X) \geq 0$.
We note that $\lambda(X) \overset{D}{\to} W \sim \chi^2(k - q)$ under H_0 where k is the number of unknown parameters in θ and q is the number of unknown parameters in θ *under* H_0.

35.2 LRT example with composite Gaussian hypothesis testing

Example 35.1. Suppose $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ and $\sigma > 0$.
Suppose we want to test $H_0 : \mu = \mu_0$ (thus $H_a : \mu \neq \mu_0$).

Solution. Since we want to test just $\mu = \mu_0$ (for any arbitrary σ), this is a composite null hypothesis. Note that $\Omega = \{(\mu, \sigma) \mid \mu \in \mathbb{R}, \sigma > 0\}$ (thus $k = 2$) and $\Omega_0 = \{(\mu, \sigma) \mid \mu = \mu_0, \sigma > 0\}$ (thus $p = 1$; σ is unknown).
Thus we have

$$\lambda(X) = 2 \log \frac{L(\hat{\mu}, \hat{\sigma}, X)}{L(\tilde{\mu}, \tilde{\sigma}; X)} \xrightarrow{D} W \sim \chi^2(2-1)$$

under H_0.
We need to find $(\tilde{\mu}, \tilde{\sigma})$ and $(\tilde{\mu}, \tilde{\sigma})$ the unconstrained and constrained MLE of (μ, σ), respectively.
Recall that $\hat{\mu} = \bar{X}$ and $\hat{\sigma} = \sqrt{1/n \sum_{i=1}^n (X_i - \bar{X})^2}$.
Under H_0, $\mu = \mu_0$ therefore $\tilde{\mu} = \mu_0$. Using MLE with $\tilde{\mu} = \mu_0$ we note that $\tilde{\sigma} = \sqrt{1/n \sum_{i=1}^n (X_i - \mu_0)^2}$ (exercise: verify using 1st/2nd derivative test).
Therefore

$$\begin{aligned}
\lambda(X) &= 2\big(l(\hat{\mu}, \hat{\sigma}; X) - l(\tilde{\mu}, \tilde{\sigma}; X)\big) \\
&= 2\big(-n/2 \log(2\pi) - n \log \hat{\sigma} - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (X_i - \hat{\mu})^2 - \big(-n/2 \log(2\pi) - n \log \hat{\sigma} - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (X_i - \hat{\mu})^2\big)\big)
\end{aligned}$$

Notice that

$$\frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (X_i - \hat{\mu})^2 = \frac{1}{2} \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{1/n \sum_{i=1}^n (X_i - \bar{X})^2} = \frac{n}{2}$$

and similarly for $\tilde{\sigma}$, thus

$$\lambda(X) = 2(-\log \hat{\sigma} + n \log \tilde{\sigma}) = 2n \log \big(\frac{\tilde{\sigma}}{\hat{\sigma}}\big)$$

where $\lambda(X) \xrightarrow{D} W = \chi^2(1)$.

35.3 LRT example with two parameters from two distributions

Example 35.2. Suppose $X_i \overset{iid}{\sim} EXP(\theta_1)$ and $Y_i \overset{iid}{\sim} EXP(\theta_2)$.
Suppose we'd like to test $H_0 : \theta_1 = \theta_2$ and $H_a : \theta_1 \neq \theta_2$.

Solution. Note that $\Omega = \{(\theta_1, \theta_2) \mid \theta_1, \theta_2 > 0\}$ $(k = 2)$ and $\Omega_0 = \{(\theta_1, \theta_2) \mid \theta_1 = \theta_2, \theta_2 > 0\}$ $(q = 1)$.
Therefore $\lambda(X) \sim \chi^2(2-1)$ under H_0.
Note that

$$\begin{aligned}
L(\theta_1, \theta_2) &= P(X_1 = x_1, \dots, X_n = x_n, Y_1 = y_1, \dots, Y_n = y_n) \\
&= \prod_{i=1}^n P(X_i = x_i) P(Y_i = y_i) && \text{independence} \\
&= \big(\prod_{i=1}^n \frac{1}{\theta_1} e^{-X_i/\theta_1}\big)\big(\prod_{i=1}^n \frac{1}{\theta_2} e^{-Y_i/\theta_2}\big)
\end{aligned}$$

therefore taking the log

$$l(\theta_1, \theta_2) = -n \log \theta_1 - \frac{1}{\theta_1} \sum_{i=1}^n X_i - n \log \theta_2 - \frac{1}{\theta_2} \sum_{i=1}^n Y_i$$

solving for $\frac{\partial l(\theta_1, \theta_2)}{\partial \theta_1} = 0$ and $\frac{\partial l(\theta_1, \theta_2)}{\partial \theta_2} = 0$ simultaneously gives us $\hat{\theta_1} = \bar{X}$ and $\hat{\theta_2} = \bar{Y}$.

We verify that

$$I(\hat{\theta_1}, \hat{\theta_2}) = \begin{bmatrix} n/\bar{X}^2 & 0 \\ 0 & n/\bar{Y}^2 \end{bmatrix}$$

is positive definite to ensure we indeed have the MLE.
Secondly, under H_0 we have $\theta_1 = \theta_2$ so

$$l(\theta_1, \theta_1) = -2n \log \theta_1 - \frac{1}{\theta_1}(\sum_{i=1}^{n} X_i + Y_i)$$

solving for $\frac{\partial l(\theta_1,\theta_1)}{\partial \theta_1} = 0$ and $\frac{\partial l(\theta_1,\theta_1)}{\partial \theta_2} = 0$ simultaneously gives us $\tilde{\theta}_1 = \frac{\bar{X}+\bar{Y}}{2} = \tilde{\theta}_2$.
Therefore

$$\begin{aligned}\lambda(X) &= 2\big(l(\hat{\theta}_1, \hat{\theta}_2, X) - l(\tilde{\theta}_1, \tilde{\theta}_2, X)\big) \\ &= 2\big(- n \log \hat{\theta}_1 - n \log \hat{\theta}_2 - \frac{1}{\hat{\theta}_1} \sum_{i=1}^{n} X_i - \frac{1}{\hat{\theta}_2} \sum_{i=1}^{n} Y_i - \big(- 2n \log \theta_1 - \frac{1}{\theta_1}(\sum_{i=1}^{n} X_i + Y_i)\big)\big)\end{aligned}$$

notice that $\frac{1}{\hat{\theta}_1} \sum X_i = \frac{1}{\bar{X}} n\bar{X} = n$ and similarly $\frac{1}{\hat{\theta}_2} \sum Y_i$ and $\frac{1}{\theta_1}(\sum_{i=1}^{n} X_i + Y_i) = 2n$ so

$$\lambda(X) = 2n \log \big(\frac{(\bar{X} + \bar{Y})^2}{4\bar{X}\bar{Y}}\big)$$

For $n = 30, \sum x_i = 45, \sum y_i = 60$ we have $\lambda(x) = 0.412$ therefore the p-value $\approx P(W \geq 0.412) > P(W \geq 3.84) = 0.05$.
Therefore we do not reject $H_0 : \theta_1 = \theta_2$.