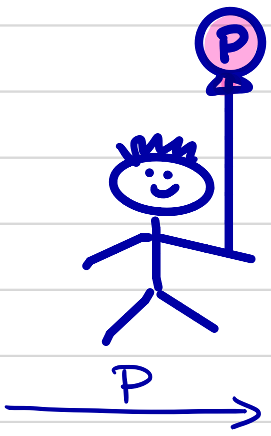
§ 5.5 Convergence Concepts



Definition 5.5.1 A sequence of random variables, $X_1, X_2, \ldots$, *converges in probability* to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} \underbrace{P(|X_n - X| \geq \epsilon)}_{\text{probability}} = 0 \quad \text{or, equivalently,} \quad \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1.$$

start here

WLLN

$\bar{X}_n \xrightarrow{P} E[X]$

Theorem 5.5.2 (Weak Law of Large Numbers) *Let $X_1, X_2, \ldots$ be iid random variables with $EX_i = \mu$ and $\operatorname{Var} X_i = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Then, for every $\epsilon > 0$,*

proof by Cheby's

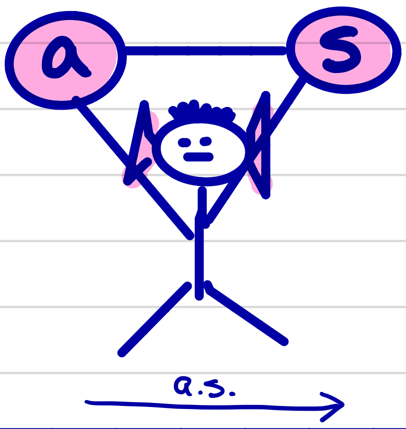
$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1;$$

that is, $\bar{X}_n$ converges in probability to μ.

If $X_n \xrightarrow{P} X$ then $h(X_n) \xrightarrow{P} h(X)$

Theorem 5.5.4 *Suppose that $X_1, X_2, \ldots$ converges in probability to a random variable X and that h is a continuous function. Then $h(X_1), h(X_2), \ldots$ converges in probability to $h(X)$.*

If $S_n^2 \xrightarrow{P} \sigma^2$ then $S_n \xrightarrow{P} \sigma$



Definition 5.5.6 A sequence of random variables, $X_1, X_2, \ldots$, *converges almost surely* to a random variable X if, for every $\epsilon > 0$,

$$P(\underbrace{\lim_{n\to\infty} |X_n - X| < \epsilon}_{\text{pointwise convergence}}) = 1.$$

start here

SLLN

$\bar{X}_n \xrightarrow{a.s.} E[X]$

Theorem 5.5.9 (Strong Law of Large Numbers) *Let $X_1, X_2, \ldots$ be iid random variables with $EX_i = \mu$ and $\operatorname{Var} X_i = \sigma^2 < \infty$, and define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Then, for every $\epsilon > 0$,*

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| < \epsilon) = 1;$$

that is, $\bar{X}_n$ converges almost surely to μ.

Definition 5.5.10 A sequence of random variables, $X_1, X_2, \ldots$, *converges in distribution* to a random variable X if

convergence in Law

$$\lim_{n\to\infty} \underline{F_{X_n}(x) = F_X(x)}$$

at all points x where $F_X(x)$ is continuous.

start here (or mgf).

CLT

$f_{\bar{X}} \xrightarrow{\mathcal{L}} N(\mu, \sigma^2/n)$

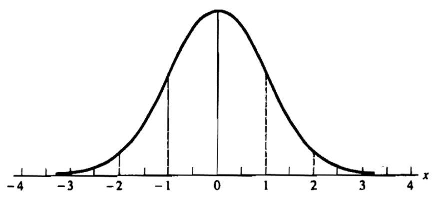


Figure 3.3.1. *Standard normal density*

Theorem 5.5.14 (Central Limit Theorem) *Let $X_1, X_2, \ldots$ be a sequence of iid random variables whose mgfs exist in a neighborhood of 0 (that is, $M_{X_i}(t)$ exists for $|t| < h$, for some positive h). Let $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\, X_i = \sigma^2 > 0$. (Both μ and σ^2 are finite since the mgf exists.) Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,*

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution.

$$\therefore \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{\mathcal{L}} N(0,1)$$

proof: homework-2

Theorem 5.5.15 (Stronger form of the Central Limit Theorem) *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $\mathrm{E}X_i = \mu$ and $0 < \mathrm{Var}\, X_i = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,*

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution.

proof uses characteristic ft'n $E[e^{itX}]$ $(i^2 = -1)$

Stronger form of CLT.

- Note no requirement for mgf to exist for $|t| < h$
- Proof similar to previous proof
  - \+ uses characteristic ft'n $E[e^{itX}]$, which always exists $\quad i^2 = -1$

    Fourier transform vs. Laplace transform
  - \+ Characteristic ft'n always exists.

\+ CLT summary

$X_1, X_2 \ldots$ iid with $E[X_i] = \mu$

$Var[X_i] = \sigma^2 < \infty$

as $n \to \infty$

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

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,1)$$

Example (Normal approximation to negative binomial)

$X_1, \ldots, X_n$ iid neg bin (r, p). [# failures to r^{th} success.]

Negative binomial(r, p)

pmf	$P(X = x \mid r, p) = \binom{r+x-1}{x} p^r (1-p)^x; \quad x = 0, 1, \ldots; \quad 0 \le p \le 1$
mean and variance	$\mathrm{E}X = \frac{r(1-p)}{p}, \quad \mathrm{Var}\, X = \frac{r(1-p)}{p^2}$
mgf	$M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r, \quad t < -\log(1-p)$
notes	An alternate form of the pmf is given by $P(Y = y \mid r, p) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$, $y = r, r+1, \ldots$. The random variable $Y = X + r$. The negative binomial can be derived as a gamma mixture of Poissons. (See Exercise 4.34.)

$E[X] = r(1-p)/p \qquad Var[X] = r(1-p)/p^2$

by CLT as $n \to \infty$ $\quad \bar{X} \sim N\left[\frac{r(1-p)}{p}, \frac{r(1-p)}{p^2}\right]$

$$\frac{\bar{X} - r(1-p)/p}{\sqrt{\frac{r(1-p)}{p^2}}/\sqrt{n}} \sim N(0,1)$$

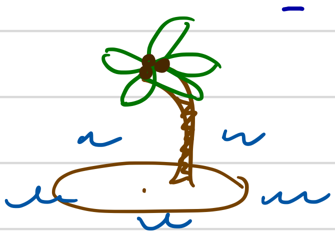
$$\frac{\sqrt{n}\,(\bar{X} - r(1-p)/p)}{\sqrt{r(1-p)/p^2}} \sim N(0,1)$$

- Assume $n=30$ trials with $r=10$, $p=1/2$

- Find $\Pr(\bar{X} < 11) = P\left(\sum_{i=1}^{30} X_i \le 330\right)$

[use mgf to show $\Sigma X_i \sim$ neg bin (nr, p), $M_{\Sigma X_i}(t) = [M_X(t)]^n$]

$$= \sum_{x=0}^{330} \binom{300+x-1}{x} \left(\tfrac{1}{2}\right)^{300} \left(\tfrac{1}{2}\right)^x$$



by hand (assuming live on small deserted island with no computer or internet) difficult calculation.

with R - not so tough...

$$\sum_{x=0}^{330} \binom{300+x-1}{x} \left(\tfrac{1}{2}\right)^{300} \left(\tfrac{1}{2}\right)^x$$

$= .8916$

Rcode: `sum(dnbinom(size=300, p=.5, x=c(0:330)))`

by CLT $P(\bar{X} \le 11) = P\left(\frac{\sqrt{30}(\bar{X}-10)}{\sqrt{20}} \le \frac{\sqrt{30}(11-10)}{\sqrt{20}}\right)$

$= P(Z \le 1.2247) = .8888$

Theorem 5.5.12 *If the sequence of random variables, $X_1, X_2, \ldots$, converges in probability to a random variable X, the sequence also converges in distribution to X.*

$$\xrightarrow{P} \overset{\text{implies}}{\Longrightarrow} \xrightarrow{\mathcal{L}}$$

Recall $\xrightarrow{a.s} \underset{\not\Leftarrow}{\Longrightarrow} \xrightarrow{P}$

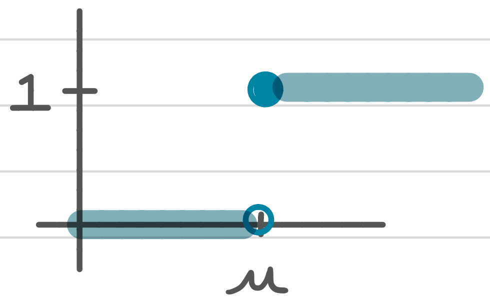
Theorem 5.5.13 *The sequence of random variables, $X_1, X_2, \ldots$, converges in probability to a constant μ if and only if the sequence also converges in distribution to μ. That is, the statement*

$$P\left(|X_n - \mu| > \varepsilon\right) \to 0 \textit{ for every } \varepsilon > 0$$

is equivalent to

$$P\left(X_n \le x\right) \to \begin{cases} 0 & \textit{if } x < \mu \\ 1 & \textit{if } x > \mu. \end{cases}$$

$X_n \xrightarrow{P} \mu \implies F_X$



Theorem 5.5.17 (Slutsky's Theorem) *If $X_n \to X$ in distribution and $Y_n \to a$, a constant, in probability, then*

a. $Y_n X_n \to aX$ *in distribution.*

b. $X_n + Y_n \to X + a$ *in distribution.*

Example: (Normal approx with estimated variance)

Suppose $\dfrac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{\mathcal{L}} N(0,1)$

But σ is unknown. Last lecture showed $S_n^2 \xrightarrow{P} \sigma^2$ (using cheby's)

$\Rightarrow S_n \to \sigma$

$\Rightarrow \sigma / S_n \to 1$

By Slutsky's Thm: $\dfrac{\sqrt{n}(\bar{X} - \mu)}{S_n} = \underbrace{\left(\dfrac{\sigma}{S_n}\right)}_{\xrightarrow{P} 1} \underbrace{\left(\dfrac{\sqrt{n}(\bar{X} - \mu)}{\sigma}\right)}_{\xrightarrow{\mathcal{L}} N(0,1)} \xrightarrow{\mathcal{L}} N(0,1)$

§ 5.5.4 Delta Method

Interested in ft'n of RV

Suppose $X_1, \dots X_n$ iid Bernoulli(p), we may be interested in estimating p.

may also be interested in estimating odds: $\left(\frac{p}{1-p}\right)$

or biostatisticians (or epidemiologists) may be interested in estimating odds ratio for two treatments: $\left(\left(p/(1-p)\right)/\left(r/(1-r)\right)\right)$

- assume estimate p by $\frac{\Sigma X_i}{n} = \frac{\#\text{ success}}{\#\text{ trials}} = \hat{p}$
- then could estimate $\frac{p}{1-p}$ by $\frac{\hat{p}}{1-\hat{p}} = \frac{\Sigma X_i/n}{1-\Sigma X_i/n}$
- How estimate $\text{Var}\left(\frac{\hat{p}}{1-\hat{p}}\right)$ or it's sampling dist'n?

"Intuition abandons us, and exact calculation is relatively hopeless, so we have to rely on an approximation."

→ Use Taylor's Expansion & Slutsky's ⟹ Delta Method

Recall: Taylor's method to approximate (mathematically) a ft'n.

Definition 5.5.20 If a function $g(x)$ has derivatives of order r, that is, $g^{(r)}(x) = \frac{d^r}{dx^r}g(x)$ exists, then for any constant a, the *Taylor polynomial of order r about a* is

$$T_r(x) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!}(x-a)^i.$$

$T_1(x) = g(a) + g'(a)(x-a)$

$T_2(x) = g(a) + g'(a)(x-a) + g''(a)(x-a)^2/2!$

Taylor's major theorem, which we will not prove here, is that the *remainder* from the approximation, $g(x) - T_r(x)$, always tends to 0 faster than the highest-order explicit term.

Theorem 5.5.21 (Taylor) *If $g^{(r)}(a) = \frac{d^r}{dx^r}g(x)\big|_{x=a}$ exists, then*

$$\lim_{x\to a}\frac{g(x) - T_r(x)}{(x-a)^r} = 0.$$

In general, we will not be concerned with the explicit form of the remainder. Since we are interested in approximations, we are just going to ignore the remainder. There are, however, many explicit forms, one useful one being

$$g(x) - T_r(x) = \int_a^x \frac{g^{(r+1)}(t)}{r!}(x-t)^r dt.$$

Use 1^{st} order Taylor's Thm $(r=1)$

Here assume $\underset{\sim}{\theta}$ is vector of length k. (Multivariate version of Taylor's).

Let $\underbrace{T_1,\dots,T_k}_{\underset{\sim}{T}}$ be RVs with means $\underbrace{\theta_1,\dots,\theta_k}_{\underset{\sim}{\theta}}$ $[E[T_i]=\theta_i \quad i=1,\dots,k]$

- Assume $g(\underset{\sim}{T})$ is differentiable ft'n (an estimator of some parameter) & we need to estimate (approximate) the variance.

- Define $$g_i'(\theta) = \frac{\partial}{\partial t_i} g(t)\Big|_{t_1=\theta_1,\dots,t_k=\theta_k} \qquad **$$

- First order Taylor's expansion of g about $\underset{\sim}{\theta}$:

Expectation $g(T) \approx g(\theta)$

$$g(\underset{\sim}{t}) = g(\underset{\sim}{\theta}) + \sum_{i=1}^{k} g_i'(\underset{\sim}{\theta})(t_i-\theta_i) + \text{Remainder}$$

ignore ≈ 0

$$g(t) \approx g(\theta) + \sum_{i=1}^{k} g_i(\underset{\sim}{\theta})(t_i-\theta_i)$$

and $$E[g(T)] \approx g(\theta) + \sum_{i=1}^{k} g_i(\underset{\sim}{\theta}) E[t_i-\theta_i] = g(\theta) + \underbrace{\left[\sum_{i=1}^{k} g_i(\theta)[\overset{=\theta_i}{E[t_i]} - \theta_i]\right]}_{=0}$$

$$\therefore E[g(T)] \approx g(\theta)$$

Variance Approx.

approximation since $g(\theta) \approx E[g(T)]$

$$\mathrm{Var}[g(T)] \approx E[[g(T)-g(\theta)]^2]$$
$$\approx E\left[\left(\sum_{i=1}^{k} g_i'(\theta)(T_i-\theta)\right)^2\right]$$
$$= \sum_{i=1}^{k} (g_i(\theta))^2 \mathrm{Var}_\theta(T_i) + 2\sum_{i>j} g_i(\theta) g_j(\theta) \mathrm{Cov}(T_i,T_j)$$

simpler example...

$$E\left[\left(\sum_{i=1}^{3}[X_i-\mu_i]\right)^2\right] =$$
$$E\left[[(X_1-\mu_1)+(X_2-\mu_2)+(X_3-\mu_3)]^2\right]$$
$$= E[(X_1-\mu_1)^2 + (X_2-\mu_2)^2 + (X_3-\mu_3)^2 \quad \text{> Variances}$$
$$+ 2(X_1-\mu_1)(X_2-\mu_2) + 2(X_1-\mu_1)(X_3-\mu_3) + 2(X_2-\mu_2)(X_3-\mu_3)] \quad \text{> Covariances}$$

<u>Example</u>: Recall odds $\left(\frac{p}{1-p}\right)$ & our estimator $\frac{\hat{p}}{1-\hat{p}} = \frac{\bar{x}}{1-\bar{x}}$

$$\mathrm{Var}\left[\frac{\hat{p}}{1-\hat{p}}\right] \approx (g'(p))^2 \mathrm{Var}(\hat{p})$$
$$= \left[\frac{1}{(1-p)^2}\right]^2 \frac{p(1-p)}{n} = \frac{p}{n(1-p)^3}$$

$$\mathrm{Var}\left[\frac{\hat{p}}{1-\hat{p}}\right] \approx \frac{p}{n(1-p)^3}$$

hmm. This is a ft'n of p, which we are estimating by $\bar{x}$! ...

$g'(p) =$ (** above)

$$\left(\frac{p}{1-p}\right)' = \frac{d}{d\bar{x}}\left(\frac{\bar{x}}{1-\bar{x}}\right)\Big|_{\bar{x}=p}$$
$$= \frac{1}{(1-p)^2}$$

Example (Approximate mean + variance)

- assume X is R.V. $E[X] = \mu \neq 0$ [here n=1]
- wish to estimate $g(\mu)$
- First order Taylor's:

$$g(X) = g(\mu) + g'(\mu)(X-\mu)$$

$$E[g(X)] \approx g(\mu)$$

$$Var[g(X)] \approx (g'(\mu))^2 \, Var[X]$$

Suppose: $g(\mu) = 1/\mu$

estimate $1/\mu$ by $1/X$ since $E[1/X] \approx 1/\mu$

$$Var[1/X] \approx (1/\mu)^4 \, Var(X)$$

hmm... again ft'n of unknown μ.
- also what if don't know $Var(X)$?

Theorem 5.5.24 (Delta Method) *Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \to n(0, \sigma^2)$ in distribution. For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not 0. Then*

(5.5.10) $$\sqrt{n}[g(Y_n) - g(\theta)] \to n(0, \sigma^2[g'(\theta)]^2) \text{ in distribution.}$$

Proof (Taylor's + Slutsky's).

Continuing above example (now have random sample, rather than n=1).

$E[\bar{X}] = \mu$ assuming finite $var(X_1) = \sigma^2$ $(Var(\bar{X}) = \sigma^2/n)$

by CLT: $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$

by Delta: $\sqrt{n}\left(\frac{1}{\bar{X}} - \frac{1}{\mu}\right) \xrightarrow{d} N\left(0, \left(\frac{1}{\mu}\right)^4 \sigma^2\right)$

or $Var\left(\frac{1}{\bar{X}}\right) \approx \left(\frac{1}{\mu}\right)^4 \sigma^2$

know $\bar{X} \xrightarrow{P} \mu$ WLLN
$S^2 \to \sigma^2$ Cheby's
Lecture 3

$$\Rightarrow \left(\frac{1}{\bar{X}}\right)^4 \xrightarrow{P} \left(\frac{1}{\mu}\right)^4$$

$$\left(\frac{1}{\bar{X}}\right)^4 S^2 \xrightarrow{P} \left(\frac{1}{\mu}\right)^4 \sigma^2$$

Theorem 5.5.4 *Suppose that $X_1, X_2, \ldots$ converges in probability to a random variable X and that h is a continuous function. Then $h(X_1), h(X_2), \ldots$ converges in probability to $h(X)$.*

Summary

$\sqrt{n}\left(\frac{1}{\bar{X}}-\frac{1}{\mu}\right) \xrightarrow{\mathcal{L}} N\left(0, \left(\frac{1}{\mu}\right)^4 \sigma^2\right)$

$\left(\frac{1}{\bar{X}}\right)^4 S^2 \xrightarrow{P} \left(\frac{1}{\mu}\right)^4 \sigma^2$

Convergence in Law $\xrightarrow{\mathcal{L}}$

Convergence in prob $\xrightarrow{P}$

Slutsky's

Theorem 5.5.17 (Slutsky's Theorem) *If $X_n \to X$ in distribution and $Y_n \to a$, a constant, in probability, then*

a. $Y_n X_n \to aX$ *in distribution.*

b. $X_n + Y_n \to X + a$ *in distribution.*

$$\frac{\sqrt{n}\left(\frac{1}{\bar{X}}-\frac{1}{\mu}\right)}{\left(\frac{1}{\bar{X}}\right)^2 S} = \underbrace{\frac{\sqrt{n}\left(\frac{1}{\bar{X}}-\mu\right)}{\left(\frac{1}{\mu}\right)^2 \sigma}}_{\xrightarrow{\mathcal{L}} N(0,1)} * \underbrace{\frac{\left(\frac{1}{\mu}\right)^2 \sigma}{\left(\frac{1}{\bar{X}}\right)^2 S}}_{\xrightarrow{P} 1} \xrightarrow{\mathcal{L}} N(0,1)$$

Extensions of Delta Method

If $g'(\theta)=0$ $\longrightarrow$

Theorem 5.5.26 (Second-order Delta Method) *Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n-\theta) \to \mathrm{n}(0,\sigma^2)$ in distribution. For a given function g and a specific value of θ, suppose that $g'(\theta)=0$ and $g''(\theta)$ exists and is not 0. Then*

$$n[g(Y_n)-g(\theta)] \to \sigma^2 \frac{g''(\theta)}{2}\chi_1^2 \text{ in distribution.} \tag{5.5.13}$$

If $\underset{\sim}{\theta}$ is vector

vector of length p

n patients, each with p observations.

Theorem 5.5.28 (Multivariate Delta Method) *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample with $\mathrm{E}(X_{ij}) = \mu_i$ and $\mathrm{Cov}(X_{ik}, X_{jk}) = \sigma_{ij}$. For a given function g with continuous first partial derivatives and a specific value of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$ for which $\tau^2 = \Sigma\Sigma\sigma_{ij}\frac{\partial g(\mu)}{\partial\mu_i}\cdot\frac{\partial g(\mu)}{\partial\mu_j} > 0$,*

$$\sqrt{n}[g(\bar{X}_1, \ldots, \underset{s=p\,?}{\bar{X}_s}) - g(\mu_1, \ldots, \mu_p)] \to \mathrm{n}(0, \tau^2) \text{ in distribution.}$$

If $\sigma_{ij} = \mathrm{Cov}(X_i, X_j)$

$\sigma_i^2 = \mathrm{Var}(X_i) = \mathrm{Cov}(X_i, X_i)$

$\rightarrow$ Define symmetric Variance / Covariance matrix

Var/Cov matrix symmetric $\sigma_{ij} = \sigma_{ji}$

$$\tau^2 = \underbrace{\underset{(1\times p)}{\begin{bmatrix} g'(\theta_1) \cdots g'(\theta_p)\end{bmatrix}} \underset{(p\times p)}{\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & & \ddots & \\ \sigma_{1p} & & & \sigma_p^2 \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} g_1'(\theta) \\ \vdots \\ g_p'(\theta)\end{bmatrix}}}_{1\times 1}$$

Note: This version of multivariate delta is ft'ns of $\bar{X}_1, \ldots, \bar{X}_s$.

n is sample size p is # parameters (# θs).

assumes $(\bar{X}_1, \ldots \bar{X}_p) \to MVN\left[\begin{pmatrix}\mu_1 \\ \vdots \\ \mu_p\end{pmatrix}, \underset{\text{var/cov matrix}}{\Sigma}\right]$ by CLT...

Example (Moments of a Ratio Estimator)

Assume $\bar{X}, \bar{Y}$ with $E[\bar{X}] = \mu$ $\quad E[\bar{Y}] = \mu_y$ $\quad Cov(X,Y) = \sigma_{xy}$

$Var[\bar{X}] = \sigma_x^2/n$ $\quad Var[\bar{Y}] = \sigma_y^2/n$

$Var[X] = \sigma_x^2$ $\quad Var[Y] = \sigma_y^2$

Further assume $g(\mu_x, \mu_y) = \frac{\mu_x}{\mu_y}$

$$\frac{\partial}{\partial \mu_x}\left[\frac{\mu_x}{\mu_y}\right] = \frac{1}{\mu_y}$$

$$\frac{\partial}{\partial \mu_y}\left[\frac{\mu_x}{\mu_y}\right] = \frac{-\mu_x}{\mu_y^2}$$

By Taylor's $E[X_1/Y_1] \cong \frac{\mu_x}{\mu_y}$

$$Var[X_1/Y_1] \cong \frac{1}{\mu_y^2}\sigma_x^2 + \frac{\mu_x^2}{\mu_y^4}\sigma_y^2 - 2\frac{\mu_x}{\mu_y^3}\sigma_{xy}$$

$$= \left(\frac{\mu_x}{\mu_y}\right)^2\left[\frac{\sigma_x^2}{\mu_x^2} + \frac{\sigma_y^2}{\mu_y^2} - \frac{2\sigma_{xy}}{\mu_x\mu_y}\right]$$

for $n=1$

$$= \left[\frac{1}{\mu_x}, \frac{-\mu_x}{\mu_y^2}\right]\begin{bmatrix}\sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2\end{bmatrix}\begin{bmatrix}1/\mu_x \\ -\mu_x/\mu_y^2\end{bmatrix}$$

§5.6 Generating a Random Sample - Focus on Probability Integral Transform.

Recall Last semester (Lecture-5)

One of most useful transformations:

Theorem 2.1.10 (Probability integral transformation) *Let X have continuous cdf $F_X(x)$ and define the random variable Y as $Y = F_X(X)$. Then Y is uniformly distributed on $(0,1)$, that is, $P(Y \le y) = y, 0 < y < 1$.*

X continuous w/ cdf $F_X(x)$

$Y = F_X(x)$

$Y \sim U(0,1) \quad P(Y \le y) = y, \; 0 < y < 1.$

Example: Assume $X \sim$ exponential

$f_X(x) = \frac{1}{\lambda} e^{-x/\lambda} I_{(0,\infty)}(x)$

$F_X(x) = \int_0^x \frac{1}{\lambda} e^{-t/\lambda} dt$

$= 1 - e^{-x/\lambda}$

[let $u = -t/\lambda$

$\frac{du}{dt} = -1/\lambda \quad du = -\frac{1}{\lambda} dt$

$\int_0^x \frac{1}{\lambda} e^{-t/\lambda} dt = \int -e^u du = -e^u \Big|$

$= -e^{-t/\lambda} \Big|_0^x = -e^{-x/\lambda} + 1$]

Let $Y = 1 - e^{-x/\lambda}$

Find dist'n of $Y = F_X(x)$

- Show monotonic ft'n:

$\frac{d}{dx}(1 - e^{-x/\lambda}) = \frac{e^{-x/\lambda}}{\lambda} > 0 \Rightarrow$ monotonic

[$\frac{d(-e^{-x/\lambda})}{dx} = (-)(-\frac{1}{\lambda}) e^{-x/\lambda}$]

- Find $g^{-1}(y)$: Solve for x in terms of y

$x = -\lambda \log(1-y) = g^{-1}(y)$

[$e^{-x/\lambda} = 1 - y$

$-\frac{x}{\lambda} = \log(1-y)$

$x = -\lambda \log(1-y)$]

- Determine sample Space, $\mathcal{Y}$

$0 \le y \le 1$

[If $x = 0 \quad y = 0$

$x \to \infty \quad y \to 1$]

- Calculate $\frac{d}{dy}(g^{-1}(y))$ (derivate old wrt new)

$\frac{d(g^{-1}(y))}{dy} = \frac{d}{dy}(-\lambda \log(1-y))$

[$\frac{d}{dy}(-\lambda \log(1-y))$

$= \frac{-\lambda}{(1-y)}(-1) = \frac{\lambda}{1-y}$]

- In $f_X(x)$ replace x by $g^{-1}(y)$ multiply by $\left|\frac{d}{dy} g^{-1}(y)\right|$ and identify y

pdf of Y $\Rightarrow f_Y(y) = 1 \, I_{(0,1)}(y)$

$Y \sim U(0,1)$

[$f_Y(y) = \frac{1}{\lambda} \exp\{-\cancel{\lambda} \log(1-y)/\cancel{\lambda}\} \Big| * \Big| \frac{\lambda}{1-y} \Big|$

$= \frac{1}{\lambda}(1-y) \Big| \frac{\lambda}{1-y} \Big| = 1$]

Go Backwards: Find dist'n $X = \lambda \log(1-y) = F_x^{-1}(y)$

If $y \sim U(0,1)$, Find dist'n of $X = -\lambda \log(1-y)$

switch $\underset{\text{new}}{X} = -\lambda \log(\underset{\text{old}}{1-y})$

- Show monotonic ft'n:

$$\frac{d}{dy}(-\lambda \log(1-y)) = \frac{\lambda}{1-y} > 0 \quad \text{for } 0<y<1$$

[previous page $\frac{d}{dy}(-\lambda\log(1-y)) = \frac{d}{dy} g^{-1}(y)$

- Find $g^{-1}(x)$: Solve for y in terms of x

$$y = 1 - e^{-x/\lambda}$$

[$y = F_x(x)$ previous page

- Determine sample Space, $\mathcal{X}$

$$0 \le X < \infty$$

[As expected if $y=0$ $x=0$
$y \to 1$ $x \to \infty$

- Calculate $\frac{d}{dx}(g^{-1}(x))$

$$\frac{d}{dx}(1 - e^{-x/\lambda}) = \frac{e^{-x/\lambda}}{\lambda}$$

[previous page $\frac{d}{dx} g(x)$

- In $f_y(y)$ replace y by $g^{-1}(x)$ multiply by $\left|\frac{d}{dx} g^{-1}(x)\right|$ and identify $\mathcal{X}$

$$f_x(x) = 1 * \left|\frac{e^{-x/\lambda}}{\lambda}\right| * I_{(0,\infty)}(x)$$

$X \sim$ exponential (λ)

Why so important? If you can generate a $U(0,1)$, you can generate random variables from any continuous dist'n.

To generate $\exp(\lambda)$ $(n = 10{,}000)$
generate 10,000 values of $y \sim U(0,1)$
calculate $X = -\lambda \log(1-y)$
$\Rightarrow$ 10,000 $\exp(\lambda)$.

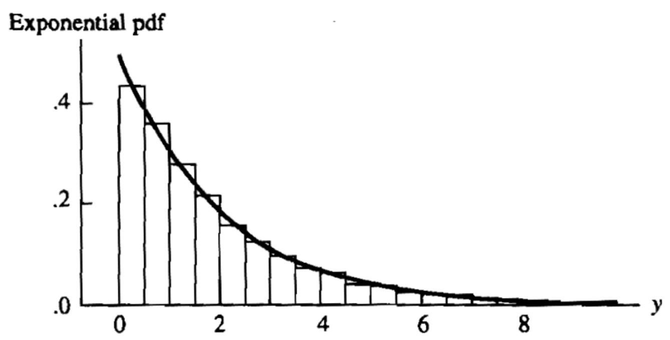


Figure 5.6.1. *Histogram of* 10,000 *observations from an exponential pdf with* $\lambda = 2$, *together with the pdf*

The relationship between the exponential and other distributions allows the quick generation of many random variables. For example, if U_j are iid uniform$(0,1)$ random variables, then $Y_j = -\lambda \log(u_j)$ are iid exponential (λ) random variables and

$$Y = -2\sum_{j=1}^{\nu} \log(U_j) \sim \chi^2_{2\nu},$$ — Note df $= 2\nu$; only generate even df

(5.6.5)
$$Y = -\beta \sum_{j=1}^{a} \log(U_j) \sim \text{gamma}(a,\beta),$$ ← Note α = integer

$$Y = \frac{\sum_{j=1}^{a} \log(U_j)}{\sum_{j=1}^{a+b} \log(U_j)} \sim \text{beta}(a,b).$$ ← Note a, b integer.

Other methods covered in C&B.

Not in C&B:

Hmm... If software will generate $N(0,1)$ how can you use this to generate a $N(\mu,\sigma^2)$?