2022

Mathematical Statistics Problem Sets w/ Solutions

MATH.UA 234

REX LIU

COURANT INSTITUTE | 251 Mercer St, NY 10012

Disclaimer:

These are the problem sets for the course Stochastic Calculus (MATH.GA 2902), given by professor Maximilian Nitzschner at New York University in Summer 2022. The solutions are mostly given by Rex Liu, with the last problem set given by Dr. Nitzschner himself.

If you see any mistakes or think that the presentation is unclear and could be improved, please send an email to: cl5682@nyu.edu. All comments and suggestions are appreciated.

Problem Set 1

Submission:
Thursday, 02/10/2022, until 1 PM, to be uploaded on the NYU Brightspace course homepage.

---

1. **Expectation and cumulative distribution function** [4 Points]

   A real random variable X is *Rayleigh-distributed* with parameter $\sigma > 0$ if its law is characterized by the probability density function

   $$f_X(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \mathbb{1}_{0,\infty)}(x).$$

   We write $X \sim Ra(\sigma)$.

   (a) Calculate the cumulative distribution function of a $Ra(\sigma)$-distributed random variable X.

   (b) Let $U \sim \mathcal{U}([0,1])$. Using the result from (a), show that for $\sigma > 0$ one has $Y = \sigma\sqrt{-2\log(U)} \sim Ra(\sigma)$.
   Hint: Calculate $F_Y(x) = \mathbf{P}[Y \leq x]$ for $x \in \mathbb{R}$.

   (c) Calculate $\mathbf{E}[X]$ and $\mathrm{Var}[X]$ for $X \sim Ra(\sigma)$.

(a) $F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} \frac{t}{\sigma^2} \exp\left(-\frac{t^2}{2\sigma^2}\right) \mathbb{1}_{[0,\infty)}(t)\,dt$

$= \int_0^x \frac{t}{\sigma^2} \exp\left(-\frac{t^2}{2\sigma^2}\right) dt$

$= -\exp\left(-\frac{t^2}{2\sigma^2}\right)\Big|_0^x$

$= \boxed{1 - e^{-\frac{x^2}{2\sigma^2}}, \quad x \in [0,\infty)}$

(b) $F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}\left(\sigma\sqrt{-2\log(U)} \leq x\right)$

$= \mathbb{P}\left(0 \leq -2\log(U) \leq \frac{x^2}{\sigma^2}\right)$

$= \mathbb{P}\left(-\frac{x^2}{2\sigma^2} \leq \log(U) \leq 0\right)$

$= \mathbb{P}\left(\exp\left(-\frac{x^2}{2\sigma^2}\right) \leq U \leq 1\right)$

$= 1 - \exp\left(-\frac{x^2}{2\sigma^2}\right)$ as $U \sim \mathcal{U}[0,1]$

$= F_X(x)$ where $X \sim Ra(\sigma)$.

So we have $Y \sim Ra(\sigma)$.

(c) $\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx$

$= \int_0^{\infty} x \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx$ integral by parts: $f = x$, $g' = x \cdot \exp\left(-\frac{x^2}{2\sigma^2}\right)$

$= \frac{1}{\sigma^2} \left[-\sigma^2 x \exp\left(-\frac{x^2}{2\sigma^2}\right) + \int \sigma^2 \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \right]_0^{\infty}$ let $u = \frac{x}{\sqrt{2}\sigma}$

$= \frac{1}{\sigma^2} \left[-\sigma^2 x \exp\left(-\frac{x^2}{2\sigma^2}\right) + \sqrt{\frac{\pi}{2}} \sigma^3 \int \frac{2e^{-u^2}}{\sqrt{\pi}} du \right]_0^{\infty}$

$= \left[-x \exp\left(-\frac{x^2}{2\sigma^2}\right) + \sqrt{\frac{\pi}{2}} \cdot \sigma \int \frac{2e^{-u^2}}{\sqrt{\pi}} du \right]_0^{\infty}$ where $u = \frac{x}{\sqrt{2}\sigma}$

$= \left[-x \exp\left(-\frac{x^2}{2\sigma^2}\right) + \sqrt{\frac{\pi}{2}} \cdot \sigma \cdot \operatorname{erf}\left(\frac{x}{\sqrt{2}\sigma}\right) \right]_0^{\infty}$

$= \sqrt{\frac{\pi}{2}} \sigma - 0 = \boxed{\sqrt{\frac{\pi}{2}} \sigma}$

$Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ where $\mathbb{E}(X)^2 = \frac{\pi \sigma^2}{2}$

$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx$

$= \int_0^{\infty} \frac{x^3}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx$ let $u = x^2$

$= \frac{1}{2\sigma^2} \int u \exp\left(-\frac{u}{2\sigma^2}\right) du$ let $v = -\frac{u}{2\sigma^2}$

$= \frac{1}{2\sigma^2} \cdot 4\sigma^4 \int v e^v dv$ integral by parts: $f = v$, $g' = e^v$.

$= 2\sigma^2 \left(v e^v - e^v \right)$

$= 2\sigma^2 \left(\frac{-u}{2\sigma^2} \exp\left(\frac{-u}{2\sigma^2}\right) - \exp\left(\frac{-u}{2\sigma^2}\right) \right)$

$= \left[-x^2 \exp\left(\frac{-x^2}{2\sigma^2}\right) - 2\sigma^2 \exp\left(\frac{-x^2}{2\sigma^2}\right) \right]_0^{\infty}$

$= 2\sigma^2 - 0 = 2\sigma^2.$

$Var(X) = \boxed{2\sigma^2 - \frac{\pi\sigma^2}{2}.}$

2. **Sums and quotients of random variables** [4 Points]

*In this problem we recall the important notion of **convolution** for two probability distributions.*

- *If X and Y are independent discrete real random variables with values in $\Omega_X, \Omega_Y \subseteq \mathbb{R}$ respectively, then $Z = X + Y$ has probability mass function*

$$p_Z(k) = \sum_{\ell \in \Omega_Y} p_X(k-\ell) p_Y(\ell),$$

for $k \in \Omega_X + \Omega_Y$, where p_X and p_Y are the probability mass functions of X and Y, and we set $p_X(r) = 0$ if $r \notin \Omega_X$.

- *If X and Y are independent continuous real random variables then $Z = X + Y$ has probability density function*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x-y) f_Y(y) \mathrm{d}y.$$

where f_X and f_Y are the probability density functions of X and Y respectively.

(a) Let $\lambda, \mu > 0$. Show that if $X \sim Pois(\lambda)$ and $Y \sim Pois(\mu)$ are independent, then $X + Y \sim Pois(\lambda + \mu)$.

(b) Let $X_1, ..., X_n$ be i.i.d. real random variables with $X_1 \sim \Gamma(\alpha, \beta)$ and $\alpha, \beta > 0$. Show that $\sum_{i=1}^n X_i \sim \Gamma(n\alpha, \beta)$.
Hint: You can use that

$$\int_0^1 u^{a-1}(1-u)^{b-1} \mathrm{d}u = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \qquad a, b > 0.$$

Proof.

(a) Let $Z = X + Y$.

Note that $P_X(x) = \mathbb{P}(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$, $P_Y(y) = \mathbb{P}(Y = y) = \frac{\mu^y}{y!} e^{-\mu}$

$$P_Z(k) = \sum_{\ell \in \Omega_Y} P_X(k-\ell) P_Y(\ell)$$

$$= \sum_{\ell=0}^{\infty} \frac{\lambda^{k-\ell}}{(k-\ell)!} e^{-\lambda} \frac{\mu^\ell}{\ell!} e^{-\mu} = \sum_{\ell=0}^{\infty} \frac{\lambda^{k-\ell}\mu^\ell}{(k-\ell)!\,\ell!} e^{-(\lambda+\mu)}$$

$$= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{\ell=0}^{\infty} \frac{k!}{(k-\ell)!\,\ell!} \lambda^{k-\ell}\mu^\ell = \frac{e^{-(\lambda+\mu)}}{k!} \sum_{\ell=0}^{\infty} \binom{k}{\ell} \lambda^{k-\ell}\mu^\ell$$

$$= \frac{(\lambda+\mu)^k}{k!} \cdot e^{-(\lambda+\mu)}, \text{ so that } Z \sim Pois(\lambda+\mu).$$

(b) first consider $Z = X_1 + X_2$.

$$f_Z(x) = \int_{-\infty}^{\infty} f_{X_1}(x-y) f_{X_2}(y) dy \quad \text{where} \quad \begin{cases} f_{X_1}(x) = \beta e^{-\beta x} \frac{(\beta x)^{\alpha-1}}{\Gamma(\alpha)}, & x \geq 0 \\ f_{X_2}(y) = \beta e^{-\beta y} \frac{(\beta y)^{\alpha-1}}{\Gamma(\alpha)}, & y \geq 0 \end{cases}$$

$$= \int_0^x \beta e^{-\beta(x-y)} \frac{(\beta(x-y))^{\alpha-1}}{\Gamma(\alpha)} \beta e^{-\beta y} \frac{(\beta y)^{\alpha-1}}{\Gamma(\alpha)} dy, \quad \text{let } u = \frac{y}{x}, \text{ then } y = ux, \text{ and } \frac{dy}{du} = x$$

$$= \beta^{2\alpha} e^{-\beta x} x^{2\alpha-1} \int_0^1 \frac{u^{\alpha-1}(1-u)^{\alpha-1}}{\Gamma(\alpha)^2} du$$

$$= \frac{\beta^{2\alpha} e^{-\beta x} x^{2\alpha-1}}{\Gamma(2\alpha)} = \beta e^{-\beta x} \frac{(\beta x)^{2\alpha-1}}{\Gamma(2\alpha)}, \text{ so that } X_1 + X_2 \sim \Gamma(2\alpha, \beta)$$

Now by induction similarly, suppose $X_1 + \cdots + X_k \sim \Gamma(k\alpha, \beta)$

we can show that $Z = (X_1 + \cdots + X_k) + X_{k+1}$ has

$$f_Z(x) = \int_{-\infty}^{\infty} f_{X_1+\cdots+X_k}(x-y) f_{X_{k+1}}(y) dy = \cdots = \beta e^{-\beta x} \frac{(\beta x)^{(k+1)\alpha-1}}{\Gamma((k+1)\alpha)}$$

Thus, $\sum_{i=1}^n X_i \sim \Gamma(n\alpha, \beta)$. ∎

(c) For the quotient of two independent, continuous, positive real random variables X and Y one can also show that $Z = \frac{X}{Y}$ has density

$$f_Z(z) = \int_0^\infty y f_X(zy) f_Y(y) \mathrm{d}y, \qquad z > 0.$$

Using this, determine the law of the quotient of two independent, $\mathcal{U}([0,1])$-distributed random variables.

(c) Let $X \sim \mathcal{U}[0,1]$ and $Y \sim \mathcal{U}[0,1]$,

so that $f_X(x) = f_Y(y) = 1$, $0 \leq x \leq 1$, $0 \leq y \leq 1$.

$\Rightarrow f_{\frac{X}{Y}}(z) = \int_0^\infty y \, f_X(zy) f_Y(y) dy$

Note that $0 \leq zy \leq 1 \Rightarrow 0 \leq y \leq \frac{1}{z}$ $(z > 0)$

for $z \geq 1$, $f_{\frac{X}{Y}}(z) = \int_0^{\frac{1}{z}} y \, dy = \frac{1}{2z^2}$

for $z < 1$, $f_{\frac{X}{Y}}(z) = \int_0^1 y \, dy = 1$

Therefore. $\boxed{f_{\frac{X}{Y}}(z) = \begin{cases} \frac{1}{2}, & 0 < z < 1 \\ \frac{1}{2z^2}, & z \geq 1 \\ 0, & \text{otherwise.} \end{cases}}$

3. **Variances and covariances** [4 Points]

(a) Let X and Y be jointly continuous with joint density

$$f_{X,Y}(x,y) = (x+y)\mathbb{1}_{\{(x,y)\in[0,1]^2\}}.$$

Calculate the covariance matrix Σ and the correlation $\rho(X,Y)$. Then use Σ to calculate $\mathrm{Var}[X-2Y]$.

(b) Let $X_1, \ldots, X_n$ be i.i.d. real random variables with $\mu = \mathbf{E}[X_1]$ and $\sigma^2 = \mathrm{Var}[X_1]$. We define

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ (sample mean)}, \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \text{ (sample variance)}.$$

Show that $\mathbf{E}[\overline{X}_n] = \mu$, $\mathrm{Var}[\overline{X}_n] = \frac{\sigma^2}{n}$ and $\mathbf{E}[S_n^2] = \sigma^2$.

(a) $f_X(x) = \int_0^1 (x+y)\,dy = xy + \frac{1}{2}y^2 \Big|_0^1 = x + \frac{1}{2}$. $\Rightarrow \mathbb{E}(X) = \int_0^1 x(x+\frac{1}{2})\,dx = \frac{7}{12}$,

$\mathbb{E}(X^2) = \int_0^1 x^2(x+\frac{1}{2})\,dx = \frac{5}{12}$.

$f_Y(y) = \int_0^1 (x+y)\,dx = \frac{1}{2}x^2 + yx \Big|_0^1 = y + \frac{1}{2}$. $\Rightarrow \mathbb{E}(Y) = \int_0^1 y(y+\frac{1}{2})\,dy = \frac{7}{12}$.

$\mathbb{E}(Y^2) = \int_0^1 y^2(y+\frac{1}{2})\,dy = \frac{5}{12}$.

So that $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{11}{144}$, $\mathrm{Var}(Y) = \frac{11}{144}$,

And $\mathbb{E}(XY) = \int_0^1\int_0^1 xy(x+y)\,dx\,dy = \frac{1}{3} = \mathbb{E}(YX)$, so that

$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = -\frac{1}{144}$

$\mathrm{Cov}(Y,X) = \mathbb{E}(YX) - \mathbb{E}(Y)\mathbb{E}(X) = -\frac{1}{144}$

$\Rightarrow \boxed{\Sigma = \begin{bmatrix} \frac{11}{144} & -\frac{1}{144} \\ -\frac{1}{144} & \frac{11}{144} \end{bmatrix}}$

$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{-\frac{1}{144}}{\frac{11}{144}} = \boxed{-\frac{1}{11}}$.

$\mathrm{Var}[X-2Y] = [1 \;\; -2]\begin{bmatrix} \frac{11}{144} & -\frac{1}{144} \\ -\frac{1}{144} & \frac{11}{144} \end{bmatrix}\begin{bmatrix} 1 \\ -2 \end{bmatrix} = \boxed{\frac{59}{144}}$

(b) Proof.

By additivity of $\mathbb{E}$ (X_i are i.i.d),

we have $\mathbb{E}(\overline{X}_n) = \mathbb{E}(\frac{1}{n}\sum_{i=1}^{n} X_n) = \frac{1}{n}\mathbb{E}(\sum_{i=1}^{n} X_n) = \frac{1}{n}\cdot n\mu = \mu$.

And since independent R.V.s are uncorrelated (i.e. $\mathrm{Cov}(X_i, X_j) = 0$)

we have $\mathrm{Var}(\overline{X}_n) = \mathrm{Var}(\frac{1}{n}\sum_{i=1}^{n} X_n) = \frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n} X_n) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n}\sigma^2$.

Now, $\mathbb{E}(S_n^2) = \mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\right)$

$= \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2)\right)$

$= \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^{n} X_i^2 - 2\overline{X}_n\left(\sum_{i=1}^{n} X_i\right) + n\overline{X}_n^2\right)$

$= \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^{n} X_i^2 - n(\overline{X}_n^2)\right)$

$= \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathbb{E}(X_i^2) - n\,\mathbb{E}(\overline{X}_n^2)\right)$, Note that $\mathbb{E}(X_i^2) = \mathrm{Var}(X_i) + \mathbb{E}(X_i)^2 = \sigma^2 + \mu^2$

and $\mathbb{E}(\overline{X}_n^2) = \mathrm{Var}(\overline{X}_n) + \mathbb{E}(\overline{X}_n)^2 = \frac{\sigma^2}{n} + \mu^2$

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

$= \frac{1}{n-1}(n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2) = \sigma^2$. $\square$

4. **Moment generating functions** [4 Points]

In this problem we recall the ***moment generating functions*** *for random variables. For a real random variable X, the moment generating function is defined by*

$$\psi_X(t) = \mathbf{E}\left[e^{tX}\right], \qquad t \in \mathbb{R},$$

whenever this expression exists. One has:

- *Whenever $\psi_X(t) = \psi_Y(t)$ for $t \in (-\varepsilon, \varepsilon)$, $\varepsilon > 0$, for two random variables X and Y, then $X \overset{d}{=} Y$.*
- $\psi'_X(0) = \mathbf{E}[X]$.

(a) Calculate the moment generating function of $X \sim \mathcal{N}(0,1)$.

(b) Suppose that X and Y are independent. Explain why $\psi_{X+Y}(t) = \psi_X(t)\psi_Y(t)$ (assuming all these expressions exist for a given t). Use this to show that if $X \sim \mathcal{N}(0, \sigma_1^2)$ and $Y \sim \mathcal{N}(0, \sigma_2^2)$, then $X + Y \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$, where $\sigma_1, \sigma_2 > 0$.

(c) The moment generating function of $X \sim Geo(p)$ with $p \in (0,1)$ is given by

$$\psi_X(t) = \frac{pe^t}{1-(1-p)e^t}, \qquad t < -\log(1-p).$$

Use this to verify that $\mathbf{E}[X] = \frac{1}{p}$.

(a) $\psi_X(t) = \mathbb{E}(e^{tX})$ where $X \sim \mathcal{N}(0,1)$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} e^{tx}\, dx$$

$$= \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2}\, dx\right) e^{\frac{t^2}{2}}$$ let $u = \frac{x-t}{\sqrt{2}} \Rightarrow du/dx = 1/\sqrt{2}$

$$= \left(\frac{1}{\sqrt{2\pi}} \cdot \left[\frac{\sqrt{\pi}}{\sqrt{2}} \operatorname{erf}(u)\right]\right) e^{\frac{t^2}{2}}$$

$$= \left(\frac{1}{\sqrt{2\pi}} \cdot \left[\frac{\sqrt{\pi}}{\sqrt{2}} \operatorname{erf}\left(\frac{x-t}{\sqrt{2}}\right)\right]_{-\infty}^{\infty}\right) e^{\frac{t^2}{2}}$$

$$= \left(\frac{\sqrt{2\pi}}{\sqrt{2\pi}}\right) e^{\frac{t^2}{2}} = \boxed{e^{\frac{t^2}{2}}}$$

(b) Proof.

$$\psi_{X+Y}(t) = \mathbb{E}\left[e^{t(X+Y)}\right] = \mathbb{E}\left[e^{tX} \cdot e^{tY}\right]$$

$$= \mathbb{E}[e^{tX}]\,\mathbb{E}[e^{tY}] = \psi_X(t)\,\psi_Y(t)$$

Note that

$$\psi_X(t) = \mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{x^2}{2\sigma_1^2}}\, dx$$

$$= e^{\frac{\sigma_1^2 t^2}{2}}$$ (similar to (a))

and $\psi_Y(t) = e^{\frac{\sigma_2^2 t^2}{2}}$.

$$\Rightarrow \psi_{X+Y}(t) = e^{\frac{\sigma_1^2 t^2}{2}} e^{\frac{\sigma_2^2 t^2}{2}} = e^{\frac{(\sigma_1^2+\sigma_2^2)}{2} t^2} = \mathbb{E}\left(e^{t(X+Y)}\right)$$

Therefore $X + Y \sim N(0, \sigma_1^2 + \sigma_2^2)$, $\sigma_1, \sigma_2 > 0$. ☒

(c) Proof. $\psi'_X(t) = \frac{d}{dt}\left(\frac{pe^t}{1-(1-p)e^t} \right)$

$$= \frac{p\left(e^t(1-(1-p)e^t) - (p-1)e^t \cdot e^t \right)}{(1-(1-p)e^t)^2}$$

for $t=0$,

$$\psi'_X(0) = \frac{p(p-p+1)}{p^2}$$

$$= \frac{1}{p}. \quad \blacksquare$$

MATH-UA.0234 Mathematical Statistics Spring 2022
Dr. M. Nitzschner

Problem Set 2

Submission:
Thursday, 02/17/2022, until 1 PM, to be uploaded on the NYU Brightspace course homepage.

---

1. **Convergence in probability and in distribution** [4 Points]

   (a) Consider a sequence $(X_n)_{n\geq 1}$ of random variables with $\mathbf{P}[X_n = n^\alpha] = \frac{1}{n}$ and $\mathbf{P}[X_n = 0] = 1 - \frac{1}{n}$ for every $n \in \mathbb{N}$. For which $\alpha \in \mathbb{R}$ does one have $X_n \xrightarrow[n\to\infty]{\mathbf{P}} 0$, and for which does one have $X_n \xrightarrow[n\to\infty]{d} 0$? For which ones does one have $\mathbf{E}[X_n] \xrightarrow[n\to\infty]{} 0$?

   (b) Suppose that $(X_n)_{n\geq 1}$ is an i.i.d. sequence of $Ber(p)$-distributed random variables. Determine the quantity a in
$$\frac{1}{n}\sum_{i=1}^{n} \exp(X_i) \xrightarrow[n\to\infty]{\mathbf{P}} a.$$

(a) $X_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$:

$\Rightarrow \lim_{n\to\infty} \mathbb{P}[\,|X_n| > \varepsilon\,] = 0, \iff \lim_{n\to\infty} \mathbb{P}(X_n > \varepsilon) = 0, \ \forall \varepsilon > 0$

Note that when $n \to \infty$, $\mathbb{P}(X_n = n^\alpha) = \frac{1}{n} \to 0$ and $\mathbb{P}(X_n = 0) = 1 - \frac{1}{n} \to 1$, $\forall \alpha \in \mathbb{R}$.

Therefore, $\boxed{\forall \alpha \in \mathbb{R}}$, we always have $\lim_{n\to\infty} \mathbb{P}(X_n > \varepsilon) = \begin{cases} \lim_{n\to\infty} \frac{1}{n} = 0 & \text{if } \varepsilon \leq n^\alpha \\ 0 & \text{if } \varepsilon > n^\alpha. \end{cases}$ $\forall \varepsilon > 0$.

$\boxed{\forall \alpha \in \mathbb{R}}$, we have $X_n \xrightarrow[n\to\infty]{d} 0$ since $X_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$, $\forall \alpha \in \mathbb{R}$. (Thm. 1.17)

$\mathbb{E}(X_n) = \frac{1}{n}\cdot n^\alpha + (1-\frac{1}{n})\cdot 0 = n^{\alpha-1}$

$\lim_{n\to\infty} \mathbb{E}(X_n) = 0$ whenever $\alpha - 1 < 0$, i.e. $\boxed{\alpha < 1}$

(b) Consider $Y_i = \exp(X_i)$:

$\mathbb{P}(Y_i = e) = \mathbb{P}(X_i = 1) = p$, $\mathbb{P}(Y_i = 1) = \mathbb{P}(X_i = 0) = 1 - p$.

$\mu = \mathbb{E}(Y_1) = ep + 1 - p$

By WLLN, we have

$\frac{1}{n}\sum_{i=1}^{n} \exp(X_i) = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{\mathbb{P}} \mu = ep + 1 - p$

So that $\boxed{a = ep + 1 - p.}$

(c) Suppose that $(X_n)_{n\geq 1}$ is an i.i.d. sequence of random variables with $\mathbf{E}[X_1] = 0$ and $\mathrm{Var}[X_1] = \sigma^2 \in (0, \infty)$. Argue in detail, why

$$\frac{1}{\log(n)\sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\mathbf{P}} 0.$$

Hint: Use the properties of convergence in probability / in distribution from the notes.

Proof.

Let $Y_n = \sqrt{n} \cdot \frac{\bar{X}_n - \mathbb{E}(X_1)}{\sigma}$ where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

By CLT, $Y_n \to Y \sim N(0,1)$.

Let $Z_n = \frac{\sigma}{\log n}$ where σ is a constant.

We have $Z_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$ as $\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{\sigma}{\log n}\right| > \varepsilon\right) = \lim_{n\to\infty} \mathbb{P}\left(\frac{\sigma}{\log n} > \varepsilon\right) \quad \forall \varepsilon > 0$

$= 0$ as $\frac{\sigma}{\log n} \to 0$ when $n \to \infty$.

Therefore, By Thm. 1.17. (iii),

$$Y_n Z_n = \sqrt{n} \cdot \frac{\bar{X}_n - \mathbb{E}(X_1)}{\sigma} \cdot \frac{\sigma}{\log n}$$

$$= \frac{\sqrt{n}}{\log n \cdot n} \sum_{i=1}^{n} X_i$$

$$= \frac{1}{\log n \sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\mathbb{P}} 0. \quad \boxtimes$$

2. **Some applications of the central limit theorem** [4 Points]

(a) Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. $\mathcal{U}(\{1,2,3,4\})$ distributed random variables. Consider the expression

$$Z_n = \log\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right).$$

Find the approximate distribution of Z_n for large n.

Hint: Use the δ-method.

(b) Suppose that the number of goals in a soccer match is Poisson-distributed with mean 3. Assume also that during a season, there are $n = 300$ matches. Use the central limit theorem to find the approximate probability for the event that during a season, there are at least 860 goals, but less than 930 goals in total. You may assume that the numbers of goals in different matches are independent.

(a) Note that $\mathbb{E}\{X_1\} = \frac{1+2+3+4}{4} = \frac{5}{2}$, $\mathrm{Var}[X_1] = \frac{5}{4} \Rightarrow \sigma = \sqrt{\frac{5}{4}} = \frac{\sqrt{5}}{2}$

Let $g(x) = \log x$, $|g'(\mathbb{E}\{X_1\})| = \frac{2}{5}$. $g(\mathbb{E}\{X_1\}) = \log\frac{5}{2}$.

By the δ- method,

$$\sqrt{n}\,\frac{g(\bar{X}_n) - g(\mu)}{|g'(\mu)|\,\sigma} = \sqrt{n}\,\frac{Z_n - \log\frac{5}{2}}{1/\sqrt{5}} \xrightarrow[n\to\infty]{d} N(0,1)$$

i.e. $\dfrac{Z_n - \log\frac{5}{2}}{1/\sqrt{5n}} \sim N(0,1)$ for large n,

$$Z_n \sim N\left(\log\frac{5}{2}, \frac{1}{5n}\right)$$

(b) Let X_i, $1 \leq i \leq 300$ be i.i.d R.V.s as number of goals in every match.

$X_1 \sim \mathrm{Poi}(3)$. i.e. $\mu = 3$, $\sigma = \sqrt{3}$.

By the CLT, we have

$$\mathbb{P}[860 \leq \Sigma_{i=1}^{300} X_i \leq 930]$$

$$= \mathbb{P}[\Sigma_{i=1}^{300} X_i < 930.5] - \mathbb{P}[\Sigma_{i=1}^{300} X_i < 859.5]$$

$$= \mathbb{P}\left[\sqrt{300}\,\frac{\bar{X}_{300} - \mu}{\sigma} < \sqrt{300}\,\frac{\frac{930.5}{300} - 3}{\sqrt{3}}\right] - \mathbb{P}\left[\sqrt{300}\,\frac{\bar{X}_{300} - \mu}{\sigma} < \sqrt{300}\,\frac{\frac{859.5}{300} - 3}{\sqrt{3}}\right]$$

$$\approx \Phi(1.02) - \Phi(-1.35)$$

$$\approx 0.846 - 0.089 \approx \boxed{0.76}$$

3. **Asymptotic normality of the t-statistics** [4 Points]

In this problem, we study the t-statistics, which is defined as follows: Let $X_1, \dots, X_n$ be i.i.d. real random variables with $\mathbf{E}[X_1] = \mu$, and $\mathrm{Var}[X_1] = \sigma^2 \in (0, \infty)$. The t-statistics is

$$T_{n-1} = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}, \qquad \textit{where } \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,\ S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

The goal of this problem is to show step-by-step that $T_{n-1} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1)$.

(a) First argue that one can write $\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ as $\sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2$.

(b) Show that $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ converges in probability to σ^2.
Hint: Apply the weak law of large numbers to both $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ and to $\overline{X}_n$. Then use the continuous mapping theorem and the fact that if $Y_n \xrightarrow[n\to\infty]{\mathbf{P}} Y$ and $Z_n \xrightarrow[n\to\infty]{\mathbf{P}} Z$, also $Y_n + Z_n \xrightarrow[n\to\infty]{\mathbf{P}} Y + Z$.

Proof.

(a) $$\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \sum_{i=1}^{n}\left(X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2\right)$$
$$= \sum_{i=1}^{n} X_i^2 - 2\overline{X}_n\sum_{i=1}^{n} X_i + \sum_{i=1}^{n}\overline{X}_n^2$$
$$= \sum_{i=1}^{n} X_i^2 - 2n\,\overline{X}_n^2 + n\,\overline{X}_n^2$$
$$= \sum_{i=1}^{n} X_i^2 - n\overline{X}_n^2$$
$$= \sum_{i=1}^{n}\left(X_i^2 - \overline{X}_n^2\right)$$

(b) $$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2 \quad (*)$$

By WLLN, $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}\{X_i^2\} = \mathrm{Var}(X_i) + \mathbb{E}(X_i)^2 = \sigma^2 + \mu^2$

and $\overline{X}_n \xrightarrow[n\to\infty]{\mathbb{P}} \mu$.

So that $(*) \xrightarrow[n\to\infty]{\mathbb{P}} \sigma^2 + \mu^2 - \mu^2 = \sigma^2$. (Thm. 1.17)

(c) Conclude from (b) that $S_n^2 \xrightarrow[n\to\infty]{\mathbf{P}} \sigma^2$. Use the continuous mapping theorem *again* to argue that $\frac{1}{S_n} \xrightarrow[n\to\infty]{\mathbf{P}} \frac{1}{\sigma}$.

(d) Finally use Slutsky's theorem and the central limit theorem to conclude that $T_{n-1} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1)$.

Hint: Note that $T_{n-1} = \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \cdot \frac{\sigma}{S_n}$.

(c) Note that $(*) \xrightarrow[n\to\infty]{\mathbb{P}} \sigma^2$ and $S_n^2 = \frac{n}{n-1}(*) \xrightarrow[n\to\infty]{\mathbb{P}} (*)$ as $\lim_{n\to\infty} \frac{n}{n-1} = 1$.

Therefore $S_n^2 \xrightarrow[n\to\infty]{\mathbb{P}} \sigma^2$, and by conti. mapping thm.,

$\frac{1}{S_n} = g(S_n^2) \xrightarrow[n\to\infty]{\mathbb{P}} g(\sigma^2) = \frac{1}{\sigma}$ where $g(x) = \frac{1}{\sqrt{x}}$.

(d) Let $Y = \sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma}$, $Z = \frac{\sigma}{S_n}$

Note that $Z \xrightarrow[n\to\infty]{\mathbb{P}} \sigma \cdot \frac{1}{\sigma} = 1$

And by CLT, $Y \xrightarrow[n\to\infty]{d} N(0,1)$.

Now, by the Hint and Slutsky's Thm, $T_{n-1} = XY \xrightarrow[n\to\infty]{d} N(0,1)$. ▨

4. **(R exercise) Simulating the central limit theorem and law of large numbers** [4 Points]

This problem illustrates the central limit theorem and the law of large numbers using simulations of random variables. You can use the software R, which can be found via *http://cran.r-project.org/*.

(a) Consider first i.i.d. random variables $X_1, ..., X_n$ with $X_1 \sim Pois(\lambda)$ with $\lambda > 0$.

(i) Formulate the law of large numbers and the central limit theorem explicitly for this situation.

(ii) To illustrate part (i), simulate $M = 5000$ times the expression in the central limit theorem that converges towards the standard normal distribution $\mathcal{N}(0,1)$, for the fixed parameter $\lambda = \frac{1}{2}$, but for different n, namely $n \in \{5, 10, 100, 1000\}$. Plot the respective histogram and the probability density function of $\mathcal{N}(0,1)$. What do you observe?

(iii) Now plot a histogram of the average $\overline{X}_n$ for $n \in \{5, 10, 100, 1000\}$ and $M = 5000$ repetitions. What do you observe?

(b) A random variable X is (standard) *Cauchy-distributed* if its law has density $f_X(x) = \frac{1}{\pi(1+x^2)}$. Repeat the simulations from (a), part (iii) but with i.i.d. Cauchy-distributed random variables $X_1, ..., X_n$. What do you observe? How can this difference be explained?

Hints:

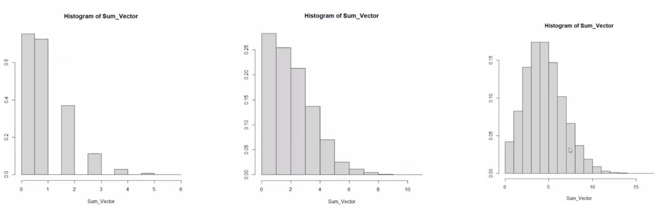
- `rpois()` generates a Poisson-distributed random variable with a chosen parameter λ.
- `rcauchy()` generates a Cauchy-distributed random variable.
- `pnorm()` gives the values of the cumulative distribution function, and `dnorm()` the values of the probability density function of a normal distribution with chosen parameters μ and σ (not σ^2).
- `hist()` plots a histogram. Use `freq =FALSE` for a normalized histogram.
- To compare the histogram to the density of a normal distribution, use the command `curve()` in the form `curve(dnorm(x,...), add=TRUE)`.
- `mean()` calculates the mean of a vector, `sqrt()` the square-root of a number.
- With `function()` you can define a new function in R.
- You can find details for every R-command by putting a ? in front for the respective function (for instance `?hist`).

(a) (i) Note that $\mu = \lambda$ and $\sigma = \sqrt{\lambda}$.

$$\text{WLLN} \Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\mathbb{P}} \lambda.$$

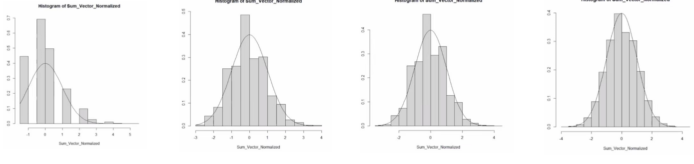
$$\text{CLT} \Rightarrow \sqrt{n}\,\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \lambda}{\sqrt{\lambda}} \xrightarrow[n\to\infty]{d} N(0,1).$$

(ii) Selected Plots:



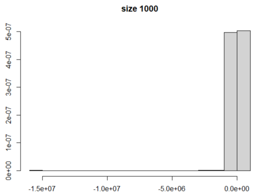
Observations: when n gets larger, the distribution is closer to the shape of the poisson distribution

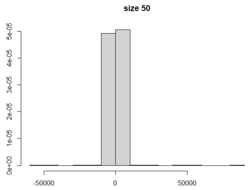
(iii) Selected Plots:



Observations: as $\overline{X}_n$ normalized the R.V.s, the distribution is closer to the standard $N(0,1)$ distribution as n gets larger.

(b) Selected Plots:





Observations: no matter what n we choose, it does not correspond with N(0, 1).

Difference explanation: The Cauchy distibution has neither a mean or variance, So that the CLT does not apply.

Problem Set 3

Submission:
Thursday, 02/24/2022, until 1 PM, to be uploaded on the NYU Brightspace course homepage.

---

1. **On the multivariate normal distribution** [4 Points]

   Let $X = (X_1, X_2)$ be a random vector with $X \sim \mathcal{N}_2(\mu, \Sigma)$, where

$$\mu = \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}.$$

   (a) Write down the probability density function of the law of X.

   (b) Find the distribution of $Y = X_1 - 3X_2$ and of $Z = (X_1 + X_2, X_1 - X_2)$.

(a)
$$f(x_1, x_2) = \frac{1}{(2\pi)^{2/2}\sqrt{\begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix}}} \exp\left(-\frac{1}{2}\left(\vec{x} - \begin{pmatrix} 1 \\ -2 \end{pmatrix}\right)^T \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \left(\vec{x} - \begin{pmatrix} 1 \\ -2 \end{pmatrix}\right)\right)$$

$$= \boxed{\frac{1}{2\sqrt{5}\,\pi} \exp\left(-\frac{2x_1^2 - 2x_1x_2 + 3x_2^2 - 8x_1 + 14x_2 + 18}{10}\right)}$$

(b) Note that $Y = [1\ -3]X$ and $Z = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} X$

By Lemma 2.2.(ii),

$\mu_Y = [1\ -3]\begin{bmatrix} 1 \\ -2 \end{bmatrix} = 7$. $\Sigma_Y = [1\ -3]\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ -3 \end{bmatrix} = 15$;

$\mu_Z = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$. $\Sigma_Z = \begin{bmatrix} 7 & 1 \\ 1 & 3 \end{bmatrix}$; so that

$$\boxed{\begin{array}{l} Y \sim \mathcal{N}(7, 15) \\ Z \sim \mathcal{N}_2(\mu, \Sigma) \text{ where } \mu = \begin{pmatrix} -1 \\ 3 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} 7 & 1 \\ 1 & 3 \end{pmatrix} \end{array}}$$

(c) Consider $U \sim \mathcal{N}(0,1)$ and $R \sim Ber(\frac{1}{2})$, independent from U. Set

$$V = \begin{cases} U, & R = 1, \\ -U, & R = 0. \end{cases}$$

It can be proved that $V \sim \mathcal{N}(0,1)$.[1] Show that $\mathrm{Cov}[U,V] = 0$, and that U and V are *not* independent. Why is this not a contradiction to Remark 2.3 (i)?

(c) Proof.

Note that $\mathbb{P}(R=1) = \mathbb{P}(R=0) = \frac{1}{2}$ and $\mathbb{E}(U) = 0$. (1)

Consider UV. $UV = \begin{cases} U^2, & R=1 \\ -U^2, & R=0 \end{cases}$

$$\mathbb{P}[UV \le x] = \frac{1}{2}\mathbb{P}[U^2 \le x] + \frac{1}{2}\mathbb{P}[U^2 \ge -x]$$

$$= \begin{cases} \frac{1}{2}\mathbb{P}[-\sqrt{x} \le U \le \sqrt{x}] + \frac{1}{2}, & x > 0 \\ \frac{1}{2}\left(\mathbb{P}[U \ge \sqrt{-x}] + \mathbb{P}[U \le -\sqrt{-x}]\right), & x \le 0 \end{cases}$$

$$= \begin{cases} \mathbb{P}[U \le -\sqrt{-x}] & x \le 0 \\ \frac{1}{2} + \mathbb{P}[0 \le U \le \sqrt{x}] = \mathbb{P}[U \le \sqrt{x}] & x > 0 \end{cases}$$

which is symmetric w/ respect to $x = 0$ as $U \sim N(0,1)$. (see also 2(a))

Thus $\mathbb{E}[XY] = 0$. (2)

$\Rightarrow \mathrm{Cov}[U,V] := \mathbb{E}[UV] - \mathbb{E}[U]\mathbb{E}[V] = 0$. (by (1) and (2))

Obviously U and V are not independent as V has an exact relationship with U, namely $V = \begin{cases} U, & R=1 \\ -U, & R=0. \end{cases}$

However it's not a contradiction to Remark 2.3.(i), which states that if X and Y are jointly normal distributed, then "pairwise uncorrolated" $\Rightarrow$ "independence".

By def., jointly normal means $aX + bY \sim$ Normal Distribution, $\forall\, a, b \in \mathbb{R}$.

However, U and V does not satisfy this def. for

$Z = U + V = \begin{cases} 2U, & R=1 \\ 0, & R=0 \end{cases}$ is not normal, since

$f_Z(z) = \frac{1}{2}\delta(z) + \frac{1}{4\sqrt{2\pi}} e^{-z^2/8}$ where δ is the Delta function.

Therefore $\nexists$ contradiction. ☒

[1] For instance, $\mathbf{P}[V \le v] = \frac{1}{2}\mathbf{P}[U \le v] + \frac{1}{2}\mathbf{P}[U \ge -v] = \mathbf{P}[U \le v]$ by symmetry.

2. **On the χ^2-distribution** [4 Points]

(a) Calculate the density of a χ^2-distribution with one degree of freedom. For this, recall if $X \sim \mathcal{N}(0,1)$, then $Z = X^2 \sim \chi_1^2$.
Hint: Look at $F_Z(z) = \mathbf{P}[X^2 \leq z]$.

(b) Verify that the χ^2-distribution with n degrees of freedom has density

$$f_{\chi_n^2}(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} e^{-\frac{x}{2}} x^{\frac{n}{2}-1} \mathbb{1}_{(0,\infty)}(x).$$

Hint: You may use without proof the fact that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Then identify the density of χ_1^2 with a Γ-distribution and use the result from Problem 2 (b) on Problem Set 1.

(a) Let $X \sim N(0,1)$, $Z = X^2$. $\Leftrightarrow$ find PDF of Z.

$$\begin{aligned} F_Z(z) &= \mathbb{P}[X^2 \leq z] \\ &= \mathbb{P}[-\sqrt{z} \leq X \leq \sqrt{z}] \\ &= \mathbb{P}[X \leq \sqrt{z}] - \mathbb{P}[X < -\sqrt{z}] \\ &= \Phi(\sqrt{z}) - \Phi(-\sqrt{z}) \end{aligned}$$

$$\Rightarrow f_Z(z) = F_Z'(z) = \frac{1}{2\sqrt{z}} \varphi(\sqrt{z}) + \frac{1}{2\sqrt{z}} \varphi(-\sqrt{z})$$

$$= \frac{1}{\sqrt{z}} \varphi(\sqrt{z}) \quad \text{where } \varphi \text{ is the pdf of a } N(0,1) \text{ R.V. (so that } \varphi(\sqrt{z}) = \varphi(-\sqrt{z}))$$

$$\Rightarrow \boxed{f_Z(z) = \frac{1}{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-z/2}, \quad z > 0.}$$

(b) Proof.

Note that in (a) $f_Z(z)$ satisfies the p.d.f. of $\Gamma(\frac{1}{2}, \frac{1}{2})$ as $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Let $Z = \sum_{i=1}^{n} Z_i$, Z_i $(1 \leq i \leq n)$ i.i.d. $Z_1 \sim \chi_1^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$

By 2 (b) in Pset 1, $Z \sim \Gamma(\frac{n}{2}, \frac{1}{2})$

i.e. $$f_Z(z) = \frac{1}{2} e^{-z/2} \frac{(\frac{z}{2})^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})} \quad (z \geq 0)$$

$$= \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} e^{-z/2} z^{\frac{n}{2}-1} \mathbb{1}_{(0,\infty)}(z). \quad \square$$

(c) Suppose the coordinates (X_1, X_2, X_3) of a particle undergoing diffusive motion can be described at some time t by i.i.d. $\mathcal{N}(0,1)$-distributed random variables. What is the probability that the particle at time t is located within a ball of radius $\frac{1}{2}$?
Hint: Use freely available online calculators for the cumulative distribution function of the χ^2 distribution.

(c) $\Leftrightarrow$ calculate $\mathbb{P}[X_1^2 + X_2^2 + X_3^2 \leq \frac{1}{4}]$

Note that $X = X_1^2 + X_2^2 + X_3^2 \sim \chi_3^2$

$\Rightarrow \mathbb{P}[X \leq 1/4] = \boxed{0.03086.}$

3. **Derivation of the density of a t-distribution** [4 Points]

In this problem, we show step-by-step that the density of the t-distribution with n degrees of freedom is given by

$$f_{t_n}(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

(a) Recall that $T = \frac{X}{\sqrt{Y/n}}$ has the t_n-distributiuon if $X \sim \mathcal{N}(0,1)$ and $Y \sim \chi_n^2$ are independent. Write down the probability density function of $\sqrt{Y/n}$.
Hint: Look at $F_{\sqrt{Y/n}}(y) = \mathbf{P}[\sqrt{Y/n} \le y]$. The density of Y is known from Problem 2 (b).

(b) For the quotient of two independent, continuous, real random variables X and Y where Y is positive, one can also show that $Z = \frac{X}{Y}$ has density

$$f_Z(z) = \int_0^\infty y f_X(zy) f_Y(y)\mathrm{d}y, z \in \mathbb{R}.$$

Use this in combination with (a) to demonstrate that

$$(\star) \qquad f_{t_n}(z) = \frac{2^{1-\frac{n}{2}}}{\sqrt{2\pi}\Gamma(\frac{n}{2})} n^{\frac{n}{2}} \int_0^\infty y^n e^{-\frac{1}{2}(n+z^2)y^2}\mathrm{d}y.$$

Proof.

(a) Let $g(Y) = \sqrt{Y/n}$

$$\Rightarrow f_{g(Y)}(y) = \frac{d}{dy} F_Y(ny^2)$$

$$= 2ny\, f_Y(ny^2) = \frac{ny}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} e^{-\frac{ny^2}{2}} (ny^2)^{\frac{n}{2}-1}, \ y \ge 0.$$

(b) Let $W = g(Y)$, $f_W(y) = \frac{ny}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} e^{-\frac{ny^2}{2}} (ny^2)^{\frac{n}{2}-1}$

$\Rightarrow T = X/W$, by the given formula, we have

$$f_{X/W}(z) = \int_0^\infty z f_X(tz) f_Z(z)\, dz, \ t \in \mathbb{R}.$$

$$= \int_0^\infty y \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2y^2}{2}} \frac{ny}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} e^{-\frac{ny^2}{2}} (ny^2)^{\frac{n}{2}-1} dy$$

$$= \frac{2^{1-\frac{n}{2}} n^{n/2}}{\sqrt{2\pi}\,\Gamma(\frac{n}{2})} \int_0^\infty y^n e^{-\frac{1}{2}(n+z^2)y^2} dy$$

(c) Recall now that the Γ-function is defined for $\alpha > 0$ by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \mathrm{d}x.$$

Validate the formula for f_{t_n} by using a substitution in the remaining integral in (b).

(c) $$f_{t_n}(z) = \frac{2^{1-\frac{n}{2}} n^{n/2}}{\sqrt{2\pi}\,\Gamma(\frac{n}{2})} \int_0^\infty y^n e^{-\frac{1}{2}(n+z^2)y^2} dy$$

where $\int_0^\infty y^n e^{-\frac{1}{2}(n+z^2)y^2} dy$ Subs. $t = \frac{1}{2}(n+z^2)y^2$

$$= \int_0^\infty \frac{y^{n-1}}{(n+z^2)} e^{-t} dt$$

$$= \int_0^\infty \frac{\left(\frac{2t}{n+z^2}\right)^{\frac{n-1}{2}}}{(n+z^2)} e^{-t} dt$$

$$= (n+z^2)^{-\frac{n+1}{2}} 2^{\frac{n-1}{2}} \int_0^\infty t^{\frac{n-1}{2}} e^{-t} dt$$

$$\Rightarrow f_{t_n}(z) = \frac{2^{1-\frac{n}{2}} n^{n/2}}{\sqrt{2\pi}\,\Gamma(\frac{n}{2})} (n+z^2)^{-\frac{n+1}{2}} 2^{\frac{n-1}{2}} \int_0^\infty t^{\frac{n-1}{2}} e^{-t} dt$$

$$= \frac{n^{n/2}}{\sqrt{\pi}\,\Gamma(\frac{n}{2})} (n+z^2)^{-\frac{n+1}{2}} \Gamma(\tfrac{n+1}{2})$$

$$= \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\,\Gamma(\frac{n}{2})} \left(1+\frac{z^2}{n}\right)^{-\frac{n+1}{2}}. \quad \boxtimes$$

4. **Properties of estimators** [4 Points] Let X_1, ..., X_n be i.i.d. real random variables with $X_1 \sim Pois(\theta)$, with $\theta > 0$. We want to estimate $\gamma = \mathbf{P}_\theta[X_1 = 0] = e^{-\theta}$ based on the data X_1, ..., X_n. We consider the two estimators for γ:

$$\widehat{\gamma}_1 = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i=0\}}, \qquad \widehat{\gamma}_2 = \exp\left(-\frac{1}{n}\sum_{i=1}^{n} X_i\right).$$

(a) Show that $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ are both consistent estimators for γ.
Hint: Use the weak law of large numbers (WLLN) and the continuous mapping theorem (CMT).

(b) Determine whether $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ are unbiased.

(c) Calculate the mean square error $\mathrm{MSE}_\theta(\widehat{\gamma}_1) = \mathbf{E}_\theta\left[(\widehat{\gamma}_1 - \gamma)^2\right]$.

(a) Proof.

Note that $\mathbb{1}\{X_i = 0\} = \begin{cases} 1 & X_i = 0, \ \mathbb{P}_\theta = e^{-\theta} \\ 0 & X_i \neq 0. \ \mathbb{P}_\theta = 1 - e^{-\theta} \end{cases} \Rightarrow \mu = e^{-\theta}$.

$\therefore \hat{\gamma}_1 \xrightarrow{\mathbb{P}_\theta} \mu = e^{-\theta}$ by WLLN.

Note that $\bar{X}_n \xrightarrow{\mathbb{P}_\theta} \mu = \theta$ by WLLN.

By CMT, we have $\hat{\gamma}_2 = \exp(-\bar{X}_n) \xrightarrow{\mathbb{P}_\theta} \exp(-\mu) = e^{-\theta}$.

Thus they're consistent. ☒

(b) We've shown that $\mathbb{E}_\theta[\hat{\gamma}_1] \overset{\text{i.i.d}}{=} \frac{1}{n} \cdot n \cdot e^{-\theta} = e^{-\theta} = \gamma$.

So that $\hat{\gamma}_1$ is unbiased.

$$\mathbb{E}_\theta[\hat{\gamma}_2] = \mathbb{E}_\theta\left[\prod_{i=1}^{n} e^{-\frac{X_i}{n}}\right]$$
$$\overset{\text{i.i.d}}{=} \left(\mathbb{E}_\theta\left[\exp\left(-\frac{X_i}{n}\right)\right]\right)^n$$
$$= \left(e^{-\theta} e^{-\frac{\theta}{n}}\right)^n \neq e^{-\theta}$$

So that $\hat{\gamma}_2$ is biased.

(c)
$$\mathrm{MSE}_\theta(\hat{\gamma}_1) = \mathrm{Var}_\theta(\hat{\gamma}_1) + \mathrm{Bias}_\theta(\hat{\gamma}_1)^2$$
$$= \mathrm{Var}_\theta(\hat{\gamma}_1)$$
$$= \mathbb{E}_\theta(\hat{\gamma}_1^2) - \mathbb{E}_\theta(\hat{\gamma}_1)^2$$
$$= \mathbb{E}_\theta\left[\frac{1}{n^2}\left(\sum_{i=1}^{n}\mathbb{1}_{\{X_i=0\}}\right)^2\right] - e^{-2\theta}$$

where
$$\mathbb{E}_\theta\left[\left(\sum_{i=1}^{n}\mathbb{1}_{\{X_i=0\}}\right)^2\right] = \mathrm{Var}\left(\sum_{i=1}^{n}\mathbb{1}_{\{X_i=0\}}\right) + \mathbb{E}\left[\sum_{i=1}^{n}\mathbb{1}_{\{X_i=0\}}\right]^2$$
$$\overset{\mathrm{Bin}(n, e^{-\theta})}{=} n e^{-\theta}(1 - e^{-\theta}) + (n e^{-\theta})^2$$
$$= n^2 e^{-2\theta} - n e^{-2\theta} + n e^{-\theta}$$

Thus,
$$\mathrm{MSE}_\theta(\hat{\gamma}_1) = \frac{1}{n^2}\left(n^2 e^{-2\theta} - n e^{-2\theta} + n e^{-\theta}\right) - e^{-2\theta}$$
$$= \boxed{-\frac{1}{n} e^{-2\theta} + \frac{1}{n} e^{-\theta}}.$$

MATH-UA.0234 | Mathematical Statistics | Spring 2022
Dr. M. Nitzschner

Problem Set 4

Submission:
Thursday, 03/03/2022, until 1 PM, to be uploaded on the NYU Brightspace course homepage.

---

1. **A brief reminder on conditional distributions** [4 Points]

   In this problem we recall the notion of ***conditional distributions*** *for two random variables X and Y. This is in particular relevant for the notion of sufficient statistics.*

   - *If X and Y are discrete random variables with values in Ω_X and Ω_Y respectively, then for any $y \in \Omega_Y$ with $\mathbf{P}[Y = y] > 0$, the* conditional distribution of X given $Y = y$ *is characterized by the conditional probability mass function*

   $$p_{X|Y=y}(x) = \frac{\mathbf{P}[X = x, Y = y]}{\mathbf{P}[Y = y]}, \qquad x \in \Omega_X.$$

   This is the distribution of X under the probability measure $\mathbf{P}[\,\cdot\,|Y = y]$.

   - *If X and Y are continuous real random variables, then for any $y \in \mathbb{R}$ with $f_Y(y) > 0$, the* conditional distribution of X given $Y = y$ *is characterized by the conditional probability density function*

   $$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \qquad x \in \mathbb{R},$$

   where $f_{X,Y}$ is the joint probability density function of (X, Y) and f_Y is the probability density function of Y.

   (a) A fair coin is tossed 4 times. Let X denote the number of times that heads comes up and $Y = 1$ if heads comes up on the first toss and $Y = 0$ otherwise. Determine the conditional distribution of X given $Y = 1$ and the conditional distribution of Y given $X = 3$.

   (b) Consider jointly continuous random variables X, Y with density

   $$f_{X,Y}(x, y) = 4ye^{-2y(x+1)}\mathbb{1}_{\{x,y>0\}}.$$

   Determine the conditional probability density function $f_{X|Y=y}$. What is the conditional distribution of X given $Y = y$?

2. **Sufficient statistics** [4 Points]

   (a) Let $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} Pois(\lambda)$ with unknown parameter $\lambda > 0$. Show that $T(\mathbf{X}) = \sum_{j=1}^n X_j$ is a sufficient statistic for λ

   (i) ...directly, using the definition of sufficiency.

   (ii) ...using the Neyman-characterization of sufficiency.

   Hint: For (i), use the result of Problem 2 (a) on Problem set 1. You need to calculate $\mathbf{P}[X_1 = x_1, ..., X_n = x_n|T(\mathbf{X}) = t]$ for all possible values of $x_1, ..., x_n, t \in \mathbb{N}_0$.

   (b) Let $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} Pareto(\lambda, a)$ with *known* $a > 0$ and unknown $\lambda > 0$, where we say that $X \sim Pareto(\lambda, a)$ if

   $$f_X(x) = \frac{\lambda a^\lambda}{x^{\lambda+1}}\mathbb{1}_{(a,\infty)}(x).$$

   Find a sufficient statistic $T(\mathbf{X}) \in \mathbb{R}$ for λ, using the Neyman-characterization.

3. **Method of moments** [4 Points]

   (a) Consider $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} \mathcal{U}([0,\theta])$ with unknown $\theta > 0$. Calculate an estimator $\widehat{\theta}_n$ for θ based on the method of moments. Check this estimator for consistency and unbiasedness.

   (b) Consider $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} Bin(N, p)$, where $\theta = (N, p) \in \mathbb{N} \times (0, 1)$ is unknown (meaning both N and p are unknown). Determine an estimator $(\widehat{N}_n, \widehat{p}_n)$ for (N, p) based on the method of moments.

   Hint: You need both $\mathbf{E}_{(N,p)}[X_1]$ and $\mathbf{E}_{(N,p)}[X_1^2]$.

4. **(R exercise) The empirical distribution** [4 Points]

   Note: Please provide your source code and images obtained with your solution.

   Suppose that $X_1, ..., X_n$ are i.i.d. real random variables such that X_1 has cumulative distribution function F. Given a realization of these random variables, we can consider the *empirical cumulative distribution function*

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}.$$

   (a) Explain why $\widehat{F}_n(x)$ fulfills the following:

      (i) For every $x \in \mathbb{R}$, one has $\mathbf{E}[\widehat{F}_n(x)] = F(x)$ and $\mathrm{Var}[\widehat{F}_n(x)] = \frac{1}{n}F(x)(1 - F(x))$.

      (ii) For every $x \in \mathbb{R}$, one has $\widehat{F}_n(x) \xrightarrow[n\to\infty]{\mathbf{P}} F(x)$.

   Hint: For any event A in some probability space, one has $\mathbf{E}[\mathbb{1}_A] = \mathbf{P}[A]$.

   (b) Use R to generate $n \in \{5, 20, 500\}$ samples of i.i.d. random variables $X_1, ..., X_n$ following a $\mathcal{U}([0, 1])$-distribution or a $\mathcal{E}(3)$-distribution. Plot the empiricial cumulative distribution function together with the graph of the cumulative distribution function F_{X_1}. What do you observe?

   Hint: The R-command `ecdf()` calculates the empirical distribution function of a vector and `plot(ecdf())` plots the respective graph.

   (c) Data on the magnitudes of earthquakes near Fiji are available from R, using the command `quakes`. [1] For help on this dataset type `?quakes`. Plot a histogram and the empirical cumulative distribution function for the *magnitudes*. Calculate the average $\overline{X}_n$ and sample variance S_n^2 for the magnitude.

   Hint: The data set `quakes` is a data frame containing information on 5 observations (i.e. a table with 5 columns). To obtain a vector *only* containing the data in column 1, use `quakes[,4]`.

   (d) Suppose it is suggested that the data for the magnitudes $X_1, ..., X_n$ can be modelled by a $\Gamma(\alpha, \beta)$ distribution. Find consistent estimators $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ for α and β, and calculate the estimates using the data from `quakes`. Plot the cumulative distribution function of $\Gamma(\widehat{\alpha}_n, \widehat{\beta}_n)$ together with the empirical distribution function of the data. What do you observe?

   Hint: Recall that for $X \sim \Gamma(\alpha, \beta)$, we have $\mathbf{E}[X] = \frac{\alpha}{\beta}$ and $\mathrm{Var}[X] = \frac{\alpha}{\beta^2}$.

[1] alternatively, you can find this data on *https://www.stat.cmu.edu/ larry/all-of-statistics/=data/fijiquakes.dat*

Homework 4

Rex Liu

March 3, 2022

1. *Solution.* (a) The conditional distribution of X given $Y = 1$ is

$$\begin{cases} p_{X|Y=1}(1) = \frac{1/16}{1/2} = 1/8 \\ p_{X|Y=1}(2) = \frac{3/16}{1/2} = 3/8 \\ p_{X|Y=1}(3) = \frac{3/16}{1/2} = 3/8 \\ p_{X|Y=1}(4) = \frac{1/16}{1/2} = 1/8 \end{cases}$$

The conditional distribution of Y given $X = 3$ is

$$\begin{cases} p_{Y|X=3}(1) = \frac{3/16}{1/4} = 3/4 \\ p_{Y|X=3}(0) = \frac{1/16}{1/4} = 1/4 \end{cases}$$

(b) Note that $f_Y(y) = F_Y'(y) = \int_0^\infty f_{X,Y}(s,y)\,ds = \left[-2e^{-2y(s+1)}\right]_0^\infty = 2e^{-2y}$, so that we have

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{4ye^{-2y(x+1)}}{2e^{-2y}} = \boxed{2ye^{-2xy}}$$

□

2. (a) *Proof.* i. By Problem 2(a) on Pset 1, we have $T(\mathbf{X}) = \sum_{j=1}^n X_j \sim \text{Pois}\,(n\lambda)$, so that

$$\begin{aligned} &\mathbb{P}[X_1 = x_1, \cdots, X_n = x_n | T(\mathbf{X}) = t] \\ =& \frac{\mathbb{P}[X_1 = x_1, \cdots, X_n = x_n \cap T(\mathbf{X}) = t]}{\mathbb{P}[T(\mathbf{X}) = t]} \\ =& \frac{\prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i}}{e^{-n\lambda} \frac{n^t \lambda^t}{t!}} = \frac{1}{n^t} \frac{t!}{\prod_{i=1}^n x_i!} \end{aligned}$$

which is independent from λ, so that it is sufficient.

ii. Note that $f_\theta(\mathbf{X}) = \prod_{i=1}^n e^{-\lambda} \dfrac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \dfrac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$. We set $T(\mathbf{X}) = \sum_{j=1}^n X_j$, and we see that

$$f_\theta(\mathbf{X}) = g_\theta(T(\mathbf{X}))h(\mathbf{X})$$

where $g_\theta(t) = e^{-n\lambda}\lambda^t$ and $h(\mathbf{X}) = \dfrac{1}{\prod_{i=1}^n x_i!}$, so that it is sufficient by the *Neyman-characterization theorem.*

□

(b) *Solution.* Note that when $x > a$, we have $f_X(x) = \frac{\lambda a^\lambda}{x^{\lambda+1}} = \frac{1}{x} \exp\{\ln \lambda + \lambda \ln a - \lambda \ln x\}$, so that

$$\begin{aligned} f_\theta(\mathbf{X}) &= \prod_{i=1}^n f_{X_i}(x_i) \\ &= \prod_{i=1}^n \frac{1}{x_i} \exp\{\ln \lambda + \lambda \ln a - \lambda \ln x_i\} \\ &= \frac{1}{\prod_{i=1}^n x_i} \prod_{i=1}^n \exp\{\ln \lambda + \lambda \ln a - \lambda \ln x_i\} \\ &= \frac{1}{\prod_{i=1}^n x_i} \exp\{n \ln \lambda + n\lambda \ln a - \lambda \sum_{i=1}^n \ln x_i\} \end{aligned}$$

where $h(\mathbf{X}) = \frac{1}{\prod_{i=1}^n x_i}$. We set $\boxed{T(\mathbf{X}) = \sum_{i=1}^n \ln x_i}$, and we have

$$g_\theta(t) = \exp\{n \ln \lambda + n\lambda \ln a - \lambda t\}$$

Hence, by the *Neyman-characterization theorem*, this $T(\mathbf{X})$ is sufficient. □

3. *Solution.* (a) Note that

$$\mathbb{E}_\theta(X) = \int_0^\theta x \, \frac{1}{\theta} \, dx = \frac{\theta}{2} = \overline{X}_n,$$

so that $\hat{\theta}_n = 2\overline{X}_n = \boxed{\frac{2}{n} \sum_{i=1}^n X_i}$. Check unbiasedness:

$$\mathbb{E}_\theta(\hat{\theta}_n) = \frac{2}{n} \sum_{i=1}^n E_\theta(X_i) = \frac{2}{n} \, n \frac{\theta}{2} = \theta$$

Note also that $\mathbb{E}_\theta(X^2) = \frac{\theta^2}{3}$, implying that

$$\mathrm{Var}_\theta(X) = \mathbb{E}_\theta(X^2) - \mathbb{E}_\theta(X)^2 = \frac{\theta^2}{12},$$

which leads us to

$$\mathrm{Var}_\theta(\hat{\theta}_n) = \frac{4}{n^2} \sum_{i=1}^n \mathrm{Var}_\theta(X_i) = \frac{4}{n^2} n \frac{\theta^2}{12} = \frac{\theta^2}{3n}$$

By Proposition 3.10, as it is unbiased, and we have $\mathrm{Var}_\theta(\hat{\theta}_n) \to 0$ as $n \to \infty$, it is consistent.

(b) We first estimate p. Note that for *Binomial* distribution, $\mathbb{E}[X] = np$, $\mathrm{Var}[X] = np(1-p)$, so that $p = 1 - \mathrm{Var}[X]/\mathbb{E}[X] = 1 - (\mathbb{E}[X^2] - \mathbb{E}[X]^2)/\mathbb{E}[X]$. Therefore, we have

$$\hat{p}_n = 1 - \frac{\frac{1}{n}\sum_{i=1}^n X_n^2 - \overline{X}_n^2}{\overline{X}_n} = \boxed{1 - \frac{\sum_{i=1}^n X_n^2}{\sum_{i=1}^n X_n} + \overline{X}_n}$$

and hence

$$\hat{N}_n = \frac{\overline{X}_n}{\hat{p}_n} = \frac{\overline{X}_n}{1 - \frac{\sum_{i=1}^n X_n^2}{\sum_{i=1}^n X_n} + \overline{X}_n} = \boxed{\frac{n\left(\overline{X}_n\right)^2}{n\overline{X}_n - \sum_{i=1}^n X_n^2 + n\left(\overline{X}_n\right)^2}}$$

$\square$

4. *Solution.* (a) i. Note that

$$\begin{aligned}\mathbb{E}[\hat{F}_n(x)] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{P}[X_i < x] && \text{Since } \mathbb{E}[\mathbb{1}_A] = \mathbb{P}[A] \\ &= \mathbb{P}[X_1 < x] = F(x) && \text{i.i.d}\end{aligned}$$

and note that $n\hat{F}_n(x)$ is the sum of n independent Bernoulli R.V.s, i.e. it is a binomial R.V. So it follows that Var$[n\hat{F}_n(x)] = nF(x)(1-F(x))$, so that

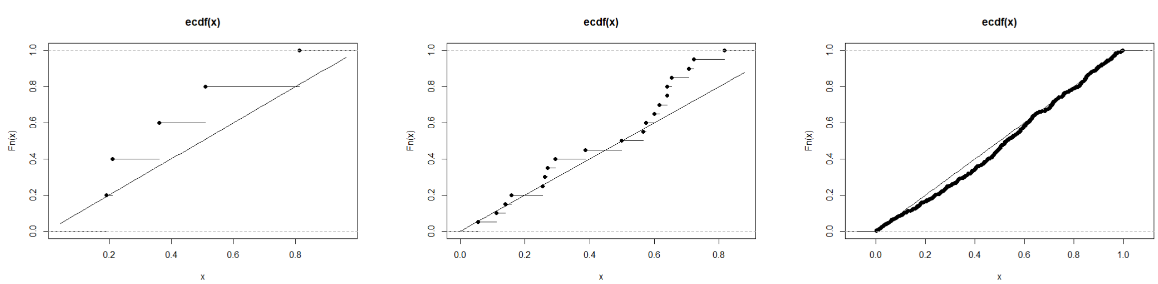
$$\begin{aligned}\text{Var}[\hat{F}_n(x)] &= \frac{1}{n^2}nF(x)(1-F(x)) \\ &= \frac{1}{n}F(x)(1-F(x))\end{aligned}$$

ii. It follows immediately by the *Chebyshev's Inequality* that for any $\epsilon > 0$, we have

$$\mathbb{P}(|F_n(x) - F(x)| \geq \epsilon) \leq \frac{F(x)(1-F(x))}{n\epsilon^2}$$

(b) Codes for the plot:

```
> x<-runif(5)
> #change 5 to 20 and 500
> hist(x)
> plot(ecdf(x))
> curve(punif, add=TRUE)
```



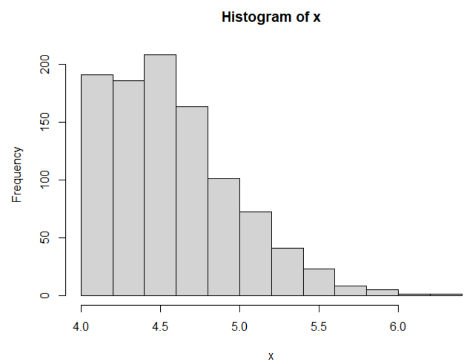
(a) Sample size 5 (b) Sample size 20 (c) Sample size 500

Figure 1: CDF and ECDF of i.i.d. R.V.s following a $\mathcal{U}([0,1])$ distribution

Observation: the more samples we pick, the closer that the *empirical cumulative distribution function* is to the actual *CDF* of X_1.

(c) Codes for the plot and calculation:

```
> x<-quakes[,4]
> hist(x)
> plot(ecdf(x))
> m<-mean(x)
> var<-var(x)
```



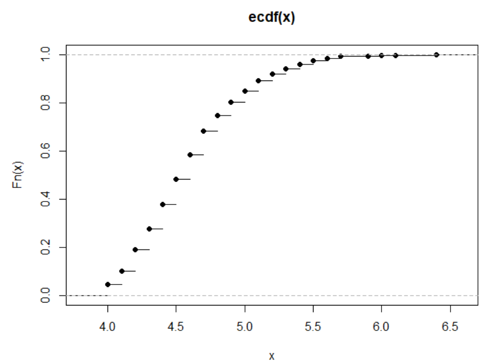


Figure 2: (left) histogram of *magnitudes*

Figure 3: (right) empirical distribution function of *magnitudes*

Calculation result: the average `m=4.6204` and sample variance `var=0.162226066066`.

(d) Since $\mathbb{E}[X] = \alpha/\beta$ and $\text{Var}[X] = \alpha/\beta^2$, we have consistent estimators $\hat{\alpha}_n = \frac{\overline{X}_n^2}{S_n^2}$ and $\hat{\beta}_n = \frac{\overline{X}_n}{S_n^2}$ where S_n^2 is the sample variance. According to the data, the estimate by solving the equations $\begin{cases} \hat{\alpha}_n/\hat{\beta}_n = 4.6204 \\ \hat{\alpha}_n/\hat{\beta}_n^2 = 0.162226066066 \end{cases}$ is $\hat{\alpha}_n = 131.5947$ and $\hat{\beta}_n = 28.48124$. Here are the codes for the plot:

```
> x<-quakes[,4]
> s_mag = sort(x)
> plot(s_mag,pgamma(s_mag, shape=131.5947,scale=1/28.48124),
   type='l')
> plot(ecdf(x),add=TRUE)
```

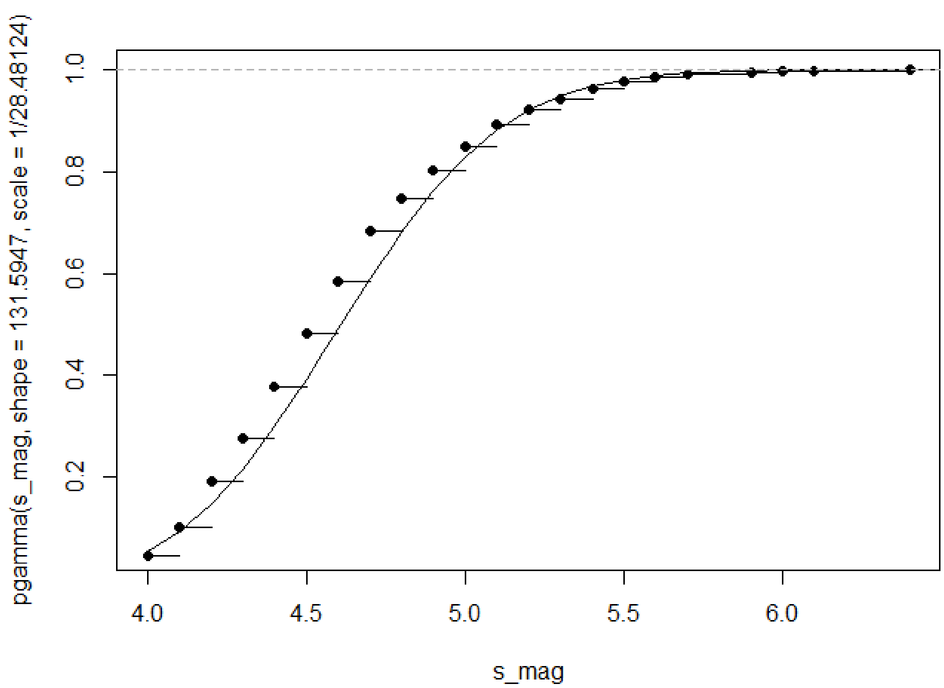


Figure 4: CDF of $\Gamma(\hat{\alpha}_n, \hat{\beta}_n)$ with given ECDF

Observation: the estimation of the CDF corresponds with the trends of the ECDF.

$\square$

Problem Set 5

Submission:
Thursday, 03/10/2022, until 1 PM, to be uploaded on the NYU Brightspace course homepage.

---

1. **Maximum-Likelihood estimator I** [4 Points]

   Let $X_1, ..., X_n$ be i.i.d. random variables and θ unknown. Calculate the Maximum-Likelihood estimator (MLE) $\widehat{\theta}_n$ for θ, if under $\mathbf{P}_\theta$,

   (a) $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, where $\theta = (\mu, \sigma^2)$,
   (b) $X_1 \sim Pois(\theta)$.
   (c) $X_1 \sim Ra(\theta)$, which means that the law of X_1 has the probability density function

   $$f(x) = \frac{x}{\theta^2}\exp\left(-\frac{x^2}{2\theta^2}\right)\mathbb{1}_{[0,\infty)}(x),$$

   with $\theta > 0$.

(a) Since $\theta = (\mu, \sigma^2)$, we let $\mu = \theta_0$, $\sigma^2 = \theta_1$

Note that $\prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta_1}} e^{-\frac{(x_i-\theta_0)^2}{2\theta_1}}$

and the log likelihood is: $\sum_{i=1}^n \left[-\log(\sqrt{2\pi\theta_1}) - \frac{1}{2\theta_1}(x_i - \theta_0)^2\right] = L(\theta)$.

$$\begin{cases} \frac{\partial L}{\partial \theta_0} = \sum_{i=1}^n \frac{x_i - \theta_0}{\theta_1} = 0 \quad \Rightarrow n\theta_0 = \sum_{i=1}^n x_i, \quad \theta_0 = \frac{1}{n}\sum_{i=1}^n x_i \\ \frac{\partial L}{\partial \theta_1} = \sum_{i=1}^n \frac{(x_i-\theta_0)^2}{2\theta_1^2} - \frac{1}{2\theta_1} = 0 \quad \Rightarrow \theta_1 = \frac{1}{n}\sum_{i=1}^n (x_i - \theta_0)^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \frac{1}{n}\sum_{i=1}^n X_i\right)^2 \end{cases}$$

$$\Rightarrow \hat{\theta}_n = \left(\overline{X}_n, \ \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2\right)$$

(b) Note that $\prod_{i=1}^n p_\theta(x_i) = \prod_{i=1}^n \frac{e^{-\theta}\theta^{x_i}}{x_i!}$

and the log likelihood is $\sum_{i=1}^n [-\theta + x_i \log\theta - \log(x_i!)] = -n\theta + \log\theta \sum_{i=1}^n x_i - \sum_{i=1}^n \log(x_i!) = L(\theta)$

$L'(\theta) = -n + \frac{1}{\theta}\sum_{i=1}^n x_i = 0$

$\Rightarrow \theta = \frac{1}{n}\sum_{i=1}^n x_i$

$$\Rightarrow \hat{\theta}_n = \overline{X}_n$$

(c) Note that $\prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \frac{x_i}{\theta^2}\exp\left(-\frac{x_i^2}{2\theta^2}\right)$ $(x \geq 0)$

and the log likelihood is $\sum_{i=1}^n \log(x_i) - 2n\log(\theta) - \frac{1}{2\theta^2}\sum_{i=1}^n x_i^2 = L(\theta)$

$L'(\theta) = \frac{-2n}{\theta} + \theta^{-3}\sum_{i=1}^n x_i^2 = 0 \Rightarrow \theta = \sqrt{\frac{1}{2n}\sum_{i=1}^n x_i^2}$ (since $\theta > 0$)

$$\Rightarrow \hat{\theta}_n = \sqrt{\frac{1}{2n}\sum_{i=1}^n X_i^2}$$

2. **Maximum-Likelihood estimator II** [4 Points]

Let $X_1, ..., X_n$ be i.i.d. random variables such that $X_1 \sim \mathcal{U}([0,\theta])$, with unknown $\theta > 0$.

(a) Calculate the Maximum-likelihood estimator (MLE) $\widehat{\theta}_n$ for θ.
Hint: Do *not* try to differentiate the (log-)likelihood function with respect to θ. Instead argue directly, for which value of θ the likelihood function is maximized.

(b) Check whether the estimator $\widehat{\theta}_n$ is consistent and unbiased.

(a) Note that $f(x|\theta) = \begin{cases} 1/\theta, & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$

So that the likelihood $L(\theta) = \begin{cases} 1/\theta^n, & 0 \le x_i \le \theta \\ 0 & \text{otherwise} \end{cases}$

Note that $\frac{1}{\theta^n}$ decreases as θ increases,

so we try to make θ as small as possible, while $\theta \ge x_i$,

$\Rightarrow$ the smallest value we can get is $\max_i x_i$, $\therefore$ $\hat{\theta}_n = \max(X_1, \dots, X_n)$.

(b) It is biased, since:

$\mathbb{P}[\hat{\theta}_n \le x] = \mathbb{P}[\cap_{i=1}^n \{x_i \le x\}] = F_X^n(x)$ (since i.i.d)

Note that $F_X(x) = \frac{1}{\theta}\int_{-\infty}^{x} \mathbb{1}_{[0,\theta]}(t)dt = \begin{cases} 0, & x<0 \\ x/\theta, & 0\le x<\theta \\ 1, & x \ge \theta \end{cases}$

therefore $f(\hat{\theta}_n) = F_X^n(x)' = \frac{n x^{n-1}}{\theta^n} \mathbb{1}_{[0,\theta]}(x)$

$\Rightarrow \mathbb{E}[\hat{\theta}_n] = \int_0^\theta x f(x) dx = \frac{n}{\theta^n}\int_0^\theta x^n dx = \frac{n}{n+1}\theta \neq \theta$

It is consistent, since:

$\mathbb{P}[\hat{\theta}_n \le x] = F_X^n(x) = \begin{cases} 0 & x<0 \\ x^n/\theta^n & 0 \le x < \theta \\ 1 & x \ge \theta \end{cases}$

When $n \to \infty$, $\mathbb{P}[\hat{\theta}_n \le x] \to \begin{cases} 0 & x<\theta \\ 1 & x \ge \theta \end{cases}$ i.e. $f(\theta)$.

Therefore, $\hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}} \theta$.

3. **Maximum-Likelihood estimator III** [4 Points]

In a pharmacological experiment, the effectiveness of an antibiotic is investigated: In n trials, dose t_1 is administered to a probe of bacterium A, then in n further trials, dose t_2 is administered. In each round it is counted, in how many of the probes the antibiotic was effective in reducing the growth of the bacteria. The probability that the growth of the bacteria is slowed by the antibiotic in a given probe is unknown, but can be modelled by $p_t = 1 - e^{-\beta t}$, where $\beta \in (0, \infty)$ is unknown.

(a) Suppose that for $j \in \{1, ..., 2n\}$,

$$X_j = \begin{cases} 1, & \text{the antibiotic reduced the growth of the bacteria in the } j\text{th trial,} \\ 0, & \text{the antibiotic did not reduce the growth of the bacteria in the } j\text{th trial.} \end{cases}$$

Write down the log-likelihood function $L_{2n}(\beta)$ for the $2n$ probes. Find an equation that must be solved by the Maximum-likelihood estimator (MLE) $\widehat{\beta}_{2n}$ for β. Which assumption(s) do you have to make in this calculation?

(a) (assumption)

Note that $X_1 \cdots X_n \sim$ i.i.d $\mathrm{Ber}(p_{t_1})$ where $p_{t_1} = 1 - e^{-\beta t_1}$

and $X_{n+1} \cdots X_{2n} \sim$ i.i.d $\mathrm{Ber}(p_{t_2})$ where $p_{t_2} = 1 - e^{-\beta t_2}$

$$\Rightarrow L(\beta) = \prod_{i=1}^{n} p_{t_1}^{x_i} (1 - p_{t_1})^{1-x_i} \prod_{i=n+1}^{2n} p_{t_2}^{x_i} (1 - p_{t_2})^{1-x_i}$$

log likelihood: $L_{2n}(\beta) = \log p_{t_1} \sum_{i=1}^{n} x_i + \log(1 - p_{t_1}) \sum_{i=1}^{n} (1 - x_i) + \log p_{t_2} \sum_{i=n+1}^{2n} x_i + \log(1 - p_{t_2}) \sum_{i=n+1}^{2n} (1 - x_i)$

$$= \log(1 - e^{-\beta t_1}) \sum_{i=1}^{n} x_i - \beta t_1 \sum_{i=1}^{n} (1 - x_i)$$
$$+ \log(1 - e^{-\beta t_2}) \sum_{i=n+1}^{2n} x_i - \beta t_2 \sum_{i=n+1}^{2n} (1 - x_i)$$

To get $\hat{\beta}_{2n}$, we let $L_{2n}(\beta)' = 0$ and find β.

(b) Suppose that there are $n = 10$ trials with dose $t_1 = 1$, and $n = 10$ further trials with dose $t_2 = 2$. A reduction of bacteria growth is observed in 3 trials of the first group, and 5 trials of the second group. Find $\widehat{\beta}_{20}$ based on the given information.

(c) Based on the model $p_t = 1 - e^{-\beta t}$, what would be an estimate of the probability be that the antibiotic reduces the growth of the bacteria when dose $t_3 = 4$ is administered?

(b) Plugging in the data, we have

$$L_{2n}(\beta) = 3\log(1-e^{-\beta}) + 5\log(1-e^{-2\beta}) - 17\beta$$

$$\Rightarrow L_{2n}(\beta)' = \frac{17e^{2\beta} - 3e^{\beta} - 30}{1-e^{2\beta}} = 0,$$ solving the equation, we get

$$\hat{\beta}_{20} = \log\left(\frac{3+\sqrt{2049}}{34}\right) \approx 0.35$$

(c) $$\hat{p} \approx 1 - e^{-\hat{\beta} t_3}$$

$$= 1 - e^{-4\times 0.35} \approx \underline{0.75}.$$

4. **Caveats for unbiased estimators** [4 Points]

In this problem, we illustrate what can "go wrong" by focussing too much on unbiasedness of estimators, by giving two perhaps surprising examples.

(a) Let $X_1, \ldots, X_n$ be i.i.d. random variables with $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma^2 \in (0, \infty)$. Recall that

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2, \qquad \text{where } \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

is an unbiased estimator for $\mathrm{Var}[X_1]$. Consider another estimator for σ^2 defined for $c \in (0, \infty)$ by

$$T_{n,c}^2 = c\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Show that $\mathrm{MSE}[T_{n,c}^2]$ is minimal for $c = \frac{1}{n+1}$. Interpret this result.

Hint: Recall from Theorem 2.7, (i), that $\frac{n-1}{\sigma^2} S_n^2 \sim \chi_{n-1}^2$. This together with the fact that $\chi_{n-1}^2 = \Gamma(\frac{n-1}{2}, \frac{1}{2})$ is useful to calculate $\mathrm{Var}[S_n^2]$.

Proof. $\mathrm{MSE}(T_{n,c}^2) = \mathbb{E}[(T_{n,c}^2 - \sigma^2)^2]$ $\qquad c\, S_n^2 \cdot (n-1)$

$= \mathbb{E}[(c(n-1)S_n^2 - \sigma^2)^2]$

$= \mathbb{E}[(c(n-1)S_n^2)^2] - \mathbb{E}[2c(n-1)S_n^2\sigma^2] + \mathbb{E}[(\sigma^2)^2]$

$= \mathrm{Var}(c(n-1)S_n^2) + (\mathbb{E}[c(n-1)S_n^2])^2 - \mathbb{E}[2c(n-1)S_n^2\sigma^2] + \sigma^4$

$= c^2\sigma^4\,\mathrm{Var}(\chi_{n-1}^2) + c^2\sigma^4(\mathbb{E}[\chi_{n-1}^2])^2 - 2c\sigma^4\,\mathbb{E}[\chi_{n-1}^2] + \sigma^4 \quad (*)$

where $\mathrm{Var}(\chi_{n-1}^2) = 2n-2$ and $\mathbb{E}[\chi_{n-1}^2] = n-1$

$\therefore (*) = \sigma^4(c^2(2n-2) + c^2(n-1)^2 - 2c(n-1) + 1)$

$= \sigma^4((n^2-1)c^2 - (2n-2)c + 1)$

take $c = -\dfrac{-(2n-2)}{2(n^2-1)}$

$= \dfrac{1}{n+1}$ as the minimal of the quadratic function. $\square$

Interpretation: it means that $\exists$ other estimators proportional to S_n^2 that may give a lower MSE.

(b) Now let $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} Ber(p)$ with unknown $p \in (0, 1)$. We will explain that

For $\gamma(p) = \frac{1}{p}$, no unbiased estimator exists.

We argue by contradiction: Suppose that there exists an unbiased estimator based on $X_1, ..., X_n$. One can show[1] that there also exists an unbiased estimator $\widehat{\gamma}_n$ that must factorize over the sufficient statistic $T(X_1, ..., X_n) = \sum_{i=1}^n X_i$, in other words, one can write

$$(\star) \qquad \widehat{\gamma}_n = f\left(\sum_{i=1}^n X_i\right)$$

for some function $f : \{0, ..., n\} \to \mathbb{R}$.

Suppose $\mathbf{E}_p[\widehat{\gamma}_n] = \gamma(p) = \frac{1}{p}$. Use $(\star)$ to argue that we must have

$$(\star\star) \qquad \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} f(k) = \frac{1}{p}.$$

Note that $(\star\star)$ implies that $\sum_{k=0}^n \binom{n}{k} p^{k+1} (1-p)^{n-k} f(k) - 1 = 0$. Since a polynomial of degree $n + 1 \in \mathbb{N}$ can only have $n + 1$ distinct zeroes, unless it is constant 0, this is a contradiction; In other words: the estimator $\widehat{\gamma}_n$ does not exist.

Proof. It is derived directly from the fact that $\sum_{i=1}^n X_i$ where X_i are i.i.d $Ber(p)$ has a $Bin(n, p)$ distribution that $\mathbb{P}(\sum_{i=1}^n X_i = k) = \binom{n}{k} p^k (1-p)^k$

So that $\mathbb{E}[f(\sum_{i=1}^n x_i)] = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^k f(k) = \frac{1}{p}$. ☒

[1] This follows from the Rao-Blackwell theorem, which we will see later in this course.

Problem Set 6

Submission:
Thursday, 04/07/2022, until 1 PM, to be uploaded on the NYU Brightspace course homepage.

---

1. **UMVU estimators** [4 Points]

   (a) Let $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} Ber(\theta)$, with unknown $\theta \in (0, 1)$. We are looking for the UMVU for $f(\theta) = \theta^3$.

   (i) Verify that
   $$\widehat{\gamma}(X_1, ..., X_n) = \mathbb{1}_{\{X_1=X_2=X_3=1\}}$$
   is an unbiased estimator for $f(\theta)$.

   (ii) Explain why $T(X_1, ..., X_n) = \sum_{i=1}^n X_i$ is a sufficient and complete statistic for θ.
   Hint: You can show completeness either directly or use the statement about exponential families from the notes.

   (iii) Calculate the UMVU estimator for $f(\theta)$, using the Lehmann-Scheffé theorem.

   (b) Now let $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} Pois(\theta)$, with unknown $\theta > 0$. An unbiased estimator for $f(\theta) = e^{-\theta}$ is given by $\widehat{\gamma}(X_1, ..., X_n) = \mathbb{1}_{\{X_1=0\}}$, and a sufficient and complete statistic for θ is given by $T(X_1, ..., X_n) = \sum_{i=1}^n X_i$. Use the Lehmann-Scheffé theorem to determine the UMVU estimator for $f(\theta)$.

2. **The Cramér-Rao lower bound I** [4 Points]

   Consider $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} Bin(N, p)$ with *known* $N \in \mathbb{N}$ and unknown $p \in (0, 1)$.

   (a) Calculate the Maximum-likelihood estimator (MLE) $\widehat{p}_n$ for p.

   (b) Determine the Fisher information $I(p)$ in p.

   (c) Check whether the variance of the MLE in (a) saturates the Cramér-Rao lower bound.

3. **The Cramér-Rao lower bound II** [4 Points]

   Consider $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ with unknown $(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$.

   (a) Determine the Fisher information matrix $I(\mu, \sigma^2)$ in (μ, σ^2).

   (b) Check whether the covariance matrix of the MLE $(\widehat{\mu}_n, S_n^2)$ for (μ, σ^2) saturates the Cramér-Rao lower bound.
   Hint: The MLE was calculated in Problem 1 (a) of Problem set 5.

4. **(R exercise) Maximum-Likelihood estimation for an autoregressive process** [4 Points]

   Note: Please provide your source code and images obtained with your solution.

   A *Gaussian autoregressive process* (more precisely a Gaussian AR(1) process) can be defined by considering random variables $X_0, X_1, X_2, ...$ with
   $$X_i = \theta X_{i-1} + \varepsilon_i, \text{ for } i = 1, 2, ... \qquad X_0 = 0.$$
   Here $\varepsilon_1, \varepsilon_2, ...$ are i.i.d. $\mathcal{N}(0, \sigma^2)$-distributed random variables. We will assume that $\sigma^2 > 0$ is *known* and $\theta \in \mathbb{R}$ is unknown.

(a) The joint probability density function of $X_1, \ldots, X_n$ is given by

$$f_{X_1,\ldots,X_n;\theta}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i - \theta x_{i-1}),$$

where $x_0 := 0$ and f is the density of a $\mathcal{N}(0, \sigma^2)$-distribution. Determine the Maximum-likelihood estimator $\widehat{\theta}_n$ (MLE) for θ.

(b) Simulate an autoregressive process with $\sigma^2 = 1$ and $\theta = 0.5$ with $n = 100$ time steps using R, and plot the outcome.

(c) Plot a histogram of $\widehat{\theta}_k$ for $k = 10$, $k = 100$ and $k = 1000$, $k = 5000$, each with $N = 500$ repetitions. What do you observe?

Homework 6, MATHUA-234

Rex Liu

April 7, 2022

1. **UMVU estimators**

   (a) i. Note that $\mathbb{P}[X_1 = 1] = \mathbb{P}[X_2 = 1] = \mathbb{P}[X_3 = 1] = \theta$, so, as they are i.i.d random variable, $\mathbb{P}[X_1 = X_2 = X_3 = 1] = \theta^3$. Hence $\mathbb{E}_\theta[\widehat{\gamma}] = \mathbb{E}_\theta\left[\mathbb{1}_{\{X_1=X_2=X_3=1\}}\right] = 1 \cdot \theta^3 + 0 \cdot (1-\theta^3) = \theta^3 = f(\theta) \implies \widehat{\gamma}$ is unbiased.

   ii. Sufficiency: note that

$$\mathbb{P}_\theta\left[X_1 = x_1, \ldots, X_n = x_n \mid \sum_{i=1}^n X_i = t\right] = \begin{cases} 0 & \sum_{i=1}^n x_i \neq t, \\ \frac{1}{\binom{n}{t}} & \sum_{i=1}^n x_i = t \end{cases}$$

   which is independent of θ for every $t \in \mathscr{T}$.
   Completeness: note that the PMF of $X_1 \sim Ber(\theta)$ is an exponential family:

$$p_\theta(x) = \theta^x(1-\theta)^{1-x}\mathbb{1}_{x\in\{0,1\}} = (1-\theta)\mathbb{1}_{x\in\{0,1\}} \exp\left[x \log\left(\frac{\theta}{1-\theta}\right)\right]$$

   Hence, by **Lemma 4.20.** (ii) and (iii), we deduce that $T(\mathbf{X})$ is complete.

   iii. By the *Lehmann-Scheffé* theorem and above, we have

$$\begin{aligned} \widehat{\gamma}^*(\mathbf{X}) &= \mathbb{E}_\theta\left[\widehat{\gamma}(\mathbf{X}) \mid T(\mathbf{X})\right] \\ &= \mathbb{E}_\theta\left[\mathbb{1}_{\{X_1=X_2=X_3=1\}} \mid \sum_{i=1}^n X_i\right] \end{aligned}$$

   Note that

$$\begin{aligned} \mathbb{P}_\theta\left[\mathbb{1}_{\{X_1=X_2=X_3=1\}} = 1 \mid \sum_{i=1}^n X_i = t\right] &= \frac{\mathbb{P}_\theta\left[\mathbb{1}_{\{X_1=X_2=X_3=1\}} = 1, \sum_{i=1}^n X_i = t\right]}{\mathbb{P}_\theta\left[\sum_{i=1}^n X_i = t\right]} \\ &= \frac{\mathbb{P}_\theta\left[\sum_{i=4}^n X_i = t-3\right] \cdot \theta^3}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\ &= \frac{\binom{n-3}{t-3}\theta^{t-3}(1-\theta)^{n-t} \cdot \theta^3}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{\binom{n-3}{t-3}}{\binom{n}{t}} \end{aligned}$$

So we have

$$\begin{aligned}
\mathbb{E}_\theta\left[\mathbb{1}_{\{X_1=X_2=X_3=1\}} \mid \sum_{i=1}^n X_i = t\right] &= 1\cdot\mathbb{P}_\theta\left[\mathbb{1}_{\{X_1=X_2=X_3=1\}} = 1 \mid \sum_{i=1}^n X_i = t\right] \\
&\quad + 0\cdot\mathbb{P}_\theta\left[\mathbb{1}_{\{X_1=X_2=X_3=1\}} = 0 \mid \sum_{i=1}^n X_i = t\right] \\
&= \frac{\binom{n-3}{t-3}}{\binom{n}{t}} \\
&= \frac{(n-3)!}{(t-3)!(n-t)!}\frac{t!(n-t)!}{n!} \\
&= \frac{t(t-1)(t-2)}{n(n-1)(n-2)}
\end{aligned}$$

Hence

$$\boxed{\widehat{\gamma}^*(\mathbf{X}) = \frac{\left(\sum_{i=1}^n X_i\right)\left(\sum_{i=1}^n X_i - 1\right)\left(\sum_{i=1}^n X_i - 2\right)}{n(n-1)(n-2)}}$$

is an UMVU estimator for $f(\theta)$.

(b) By the *Lehmann-Scheffé* theorem and since sum of Poisson r.v.s has Poisson distribution with parameter equal to sum of (the original) parameters, we have

$$\begin{aligned}
\widehat{\gamma}^*(\mathbf{X}) &= \mathbb{E}_\theta\left[\widehat{\gamma}(\mathbf{X}) \mid T(\mathbf{X})\right] \\
&= \mathbb{E}_\theta\left[\mathbb{1}_{\{X_1=0\}} \mid \sum_{i=1}^n X_i\right] \\
&= 1\cdot\mathbb{P}_\theta\left[\mathbb{1}_{\{X_1=0\}} = 1 \mid \sum_{i=1}^n X_i = t\right] + 0\cdot\mathbb{P}_\theta\left[\mathbb{1}_{\{X_1=0\}} = 0 \mid \sum_{i=1}^n X_i = t\right] \\
&= \mathbb{P}_\theta\left[X_1 = 0 \mid \sum_{i=1}^n X_i = t\right] \\
&= \frac{\mathbb{P}_\theta\left[X_1 = 0, \sum_{i=1}^n X_i = t\right]}{\mathbb{P}_\theta\left[\sum_{i=1}^n X_i = t\right]} \\
&= \frac{\mathbb{P}_\theta\left[X_1 = 0\right]\cdot\mathbb{P}_\theta\left[\sum_{i=2}^n X_i = t\right]}{\mathbb{P}_\theta\left[\sum_{i=1}^n X_i = t\right]} \\
&= \frac{e^{-\theta}\frac{((n-1)\theta)^t}{t!}e^{-(n-1)\theta}}{\frac{(n\theta)^t}{t!}e^{-n\theta}} \\
&= \left(\frac{n-1}{n}\right)^t = \boxed{\left(\frac{n-1}{n}\right)^{\sum_{i=1}^n X_i}}
\end{aligned}$$

as an UMVU estimator for $f(\theta)$.

2. **The Cramér-Rao lower bound I**

(a) Note that for binomial distribution, we have PMF

$$f_{X_i}(x) = \binom{N}{x} p^x (1-p)^{N-x}$$

So the likelihood function is (since i.i.d)

$$f_p^{(n)}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i) = p^{\sum_{i=1}^n x_i} (1-p)^{nN - \sum_{i=1}^n x_i} \prod_{i=1}^n \binom{n}{x_i}$$

Ignoring the constant coefficient, we let

$$\frac{\partial}{\partial p} p^{\sum_{i=1}^n x_i} (1-p)^{nN - \sum_{i=1}^n x_i} \stackrel{!}{=} 0$$

$$\iff S \cdot p^{S-1}(1-p)^{nN-S} = (nN - S)\, p^S (1-p)^{nN-S-1}$$

where $S = \sum_{i=1}^n x_i$, so that $S(1-p) = (nN-S)p \implies p = \frac{S}{nN}$. Hence the Maximum-likelihood estimator for p is $\boxed{\widehat{p}_n = \frac{\overline{X}_n}{N}}$.

(b) By definition, we have

$$\begin{aligned}
I(p) &= \mathbb{E}_p\left[\left(\frac{d \log f_p(x)}{dp}\right)^2\right] \\
&= \mathbb{E}_p\left[\left(\frac{x}{p} - \frac{N-x}{1-p}\right)^2\right] \\
&= \sum_{x=0}^N \left(\frac{x}{p} - \frac{N-x}{1-p}\right)^2 \binom{N}{x} p^x (1-p)^{N-x} \\
&= \sum_{x=0}^N \left(\frac{x^2 - 2Nxp + N^2p^2}{p^2(1-p)^2}\right) \binom{N}{x} p^x (1-p)^{N-x} \\
&= \frac{N^2p^2 + Np(1-p) - 2N^2p^2 + N^2p^2}{p^2(1-p)^2} \\
&= \boxed{\frac{N}{p(1-p)}}
\end{aligned}$$

(c) $\boxed{\text{It does.}}$ By properties of variance, we have $\mathrm{Var}(\widehat{p}_n) = \frac{nNp(1-p)}{n^2N^2} = \frac{1}{nI(p)}$, which saturates the Cramér-Rao lower bound.

3. **The Cramér-Rao lower bound II** (reference [1], [2], [3])

(a) For normal distribution, we have $\boldsymbol{\theta} = (\mu, \sigma^2)^\top$ and the log likelihood function

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{(x-\mu)^2}{2\sigma^2}$$

So the score vector is

$$s(\boldsymbol{\theta}) = \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu}, \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma^2}\right)^\top = \left(\frac{x-\mu}{\sigma^2}, -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2(\sigma^2)^2}\right)^\top$$

Note that $\mathbb{E}[X-\mu] = 0$ and $\mathrm{Var}(X) = \mathbb{E}\left[(X-\mu)^2\right] = \sigma^2$. Also, by the identity[1] that

$$\mathbb{E}\left[g(X)(X-\mu)\right] = \sigma^2 \mathbb{E}\left[g'(X)\right],$$

we have

$$\mathbb{E}\left[X^3\right] = 2\sigma^2\mu + \mu(\sigma^2+\mu^2) = \mu^3 + 3\mu\sigma^2$$

and

$$\mathbb{E}\left[X^4\right] = 3\sigma^2(\sigma^2+\mu^2) + \mu(\mu^3+3\mu\sigma^2) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$$

Using these properties, we can calculate the Fisher information matrix

$$\begin{aligned}
\mathbf{I}(\boldsymbol{\theta}) &= \mathbb{E}_{\boldsymbol{\theta}}\left[s(\boldsymbol{\theta}) \cdot s(\boldsymbol{\theta})^\top\right] \\
&= \mathbb{E}_{\boldsymbol{\theta}}\begin{bmatrix} \frac{(x-\mu)^2}{\sigma^4} & \left(\frac{x-\mu}{\sigma^2}\right)\left(-\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2(\sigma^2)^2}\right) \\ \left(\frac{x-\mu}{\sigma^2}\right)\left(-\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2(\sigma^2)^2}\right) & \left(-\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2(\sigma^2)^2}\right)^2 \end{bmatrix} \\
&= \boxed{\begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}}
\end{aligned}$$

(b) Note that the MLE is given by $\widehat{\mu}_n = \overline{X}_n$, $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$, and since they're unbiased, it's suffice to show that

$$\mathrm{Cov}(T(\mathbf{X})) \geq_L \frac{1}{n}\mathbf{I}(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix} \qquad \geq_L \text{ means the Loewner order}$$

Note that sample mean and variance are independent[2], so that we have

$$\begin{aligned}
\mathrm{Cov}(T(\mathbf{X})) - \frac{1}{n}\mathbf{I}(\boldsymbol{\theta})^{-1} &= \begin{bmatrix} \mathrm{Var}(\widehat{\mu}_n) - \frac{\sigma^2}{n} & 0 \\ 0 & \mathrm{Var}(\widehat{\sigma}_n^2) - \frac{2\sigma^4}{n} \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 \\ 0 & \mathrm{Var}(\widehat{\sigma}_n^2) - \frac{2\sigma^4}{n} \end{bmatrix}
\end{aligned}$$

whose *leading principal minors* have all determinants 0, so it is $\boxed{\textit{positive semidefinite}}$.

[1] Stein's Lemma.
[2] Proof.

4. **(R exercise) Maximum-Likelihood estimation for an autoregressive process**

   (a) Given that $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$, we have

$$f_{\mathbf{X};\theta}(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \theta x_{i-1})^2}{2\sigma^2}}$$

   Ignoring the constant coefficient $(2\pi\sigma^2)^{-n/2}$, we take $\log$ for the rest:

$$\begin{aligned}\ell(\boldsymbol{x}) &= -\sum_{i=1}^{n} \left[\frac{(x_i - \theta x_{i-1})^2}{2\sigma^2}\right] \\ &= -\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} x_i^2 - 2\theta \sum_{i=1}^{n} x_i x_{i-1} + \theta^2 \sum_{i=1}^{n} x_{i-1}^2\right]\end{aligned}$$

   which is maximized at

$$\theta = \frac{\sum_{i=1}^{n} x_i x_{i-1}}{\sum_{i=1}^{n} x_{i-1}^2} = \frac{\sum_{i=1}^{n-1} x_{i+1} x_i}{\sum_{i=1}^{n-1} x_i^2}$$

   since $x_0 := 0$, so that the MLE of θ is $\widehat{\theta}_n = \boxed{\frac{\sum_{i=1}^{n-1} X_{i+1} X_i}{\sum_{i=1}^{n-1} X_i^2}}$.

(b) Codes for the autoregressive process:

```
> # Generate an AR1 process of length n
> # Set up variables
> set.seed(1234)
> n <- 100
> x <- matrix(0,100,1)
> w <- rnorm(n)
> # loop to create x
> for (t in 2:n) x[t] <- 0.5 * x[t-1] + w[t]
> plot(x,type='l')
```

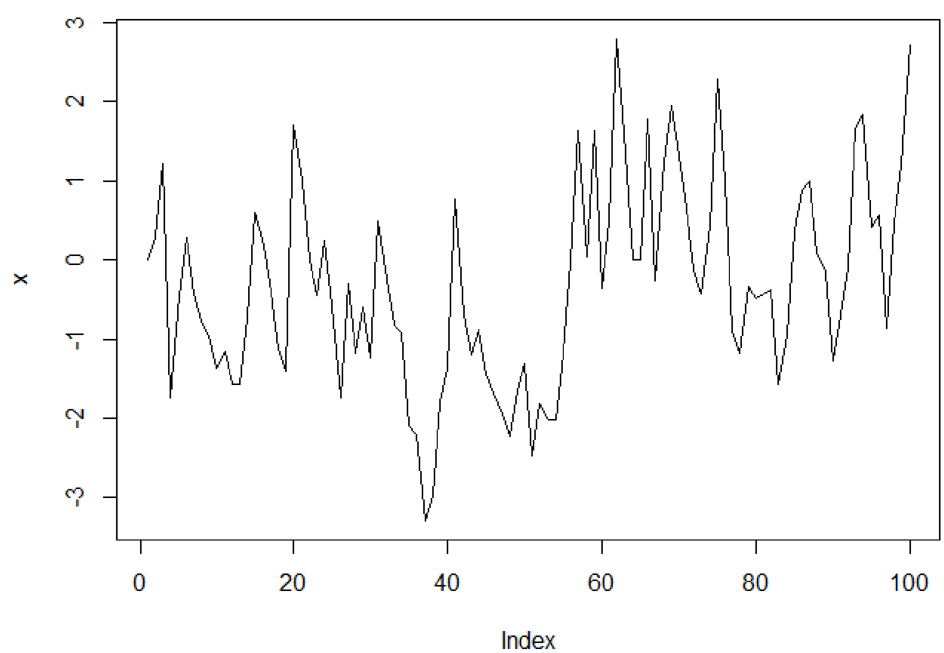
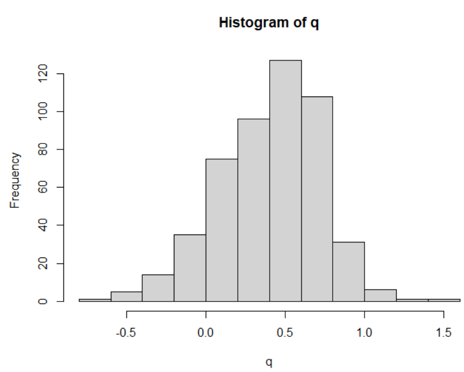


Figure 1: Outcome of the AR(1) process with $N(0,1), \theta = .5, n = 100$

(c) Observation: the larger k is, the more concentrated $\widehat{\theta}_k$ is to 0.5 (trends not obvious for small k). Codes for plotting the histogram:

```
N <- 500
k <- 10 # being changed to 100, 1000, 5000 when needed
w <- matrix( rnorm(N*k,mean=0,sd=1), N, k)
for (s in 1:N) {for (t in 2:k) w[s,t] <- 0.5 * w[s,t-1] + w[s,t]}
qu <- matrix(0,N,1)
qd <- matrix(0,N,1)
q <- matrix(0,N,1)
for (i in 1:N){for(p in 1:(k-1)){qu[i] <- w[i,p+1]*w[i,p]+qu[i]}}
for (j in 1:N){for(r in 1:(k-1)){qd[j] <- w[j,r]*w[j,r]+qd[j]}}
for (m in 1:N) q[m] = qu[m]/qd[m]
hist(q)
```



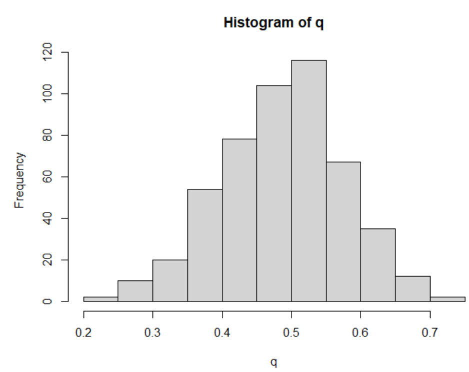
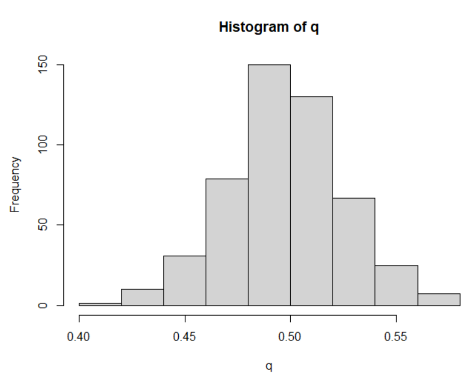


Figure 2: (left) histogram of $\widehat{\theta}_k$ for $k = 10$ w/ 500 repetitions

Figure 3: (right) histogram of $\widehat{\theta}_k$ for $k = 100$ w/ 500 repetitions



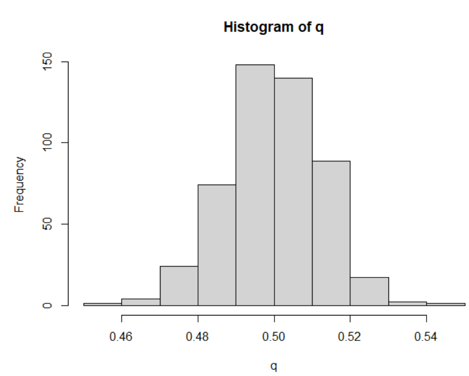


Figure 4: (left) histogram of $\widehat{\theta}_k$ for $k = 1000$ w/ 500 repetitions

Figure 5: (right) histogram of $\widehat{\theta}_k$ for $k = 5000$ w/ 500 repetitions

Homework 7, MATH.UA 234

Rex Liu

April 13, 2022

1. **Confidence intervals for normally distributed data**

   Suppose that in an experiment 5 data points are observed:

   $$x_1 = 2.3, \qquad x_2 = 1.9, \qquad x_3 = 2.0, \qquad x_4 = 1.8, \qquad x_5 = 2.1.$$

   Suppose that these data points are realizations of i.i.d. random variables $X_1, \dots, X_5$, where $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ with unknown μ and σ^2.

   (a) Find a 95%-confidence interval for μ.

   By Example 6.4, we see that

   $$S(\mathbf{X}) = \left[\overline{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1,1-\alpha/2}, \overline{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1,1-\alpha/2}\right]$$

   is a $(1-\alpha)$-confidence interval for μ. Based on the observation, we have

   $$\alpha = 0.05, \qquad n = 5, \qquad \overline{X}_n = 2.02, \qquad S_n = \sqrt{\frac{1}{4}\sum_{i=1}^{5}(X_i - 2.02)^2} \approx 0.192$$

   and $t_{4,0.975} \approx 2.776$ is the 0.975-quantile of the t_4-distribution. Hence, we have

   $$\boxed{S(\mathbf{X}) \approx \left[2.02 - \frac{0.192}{\sqrt{5}} \cdot 2.776, 2.02 + \frac{0.192}{\sqrt{5}} \cdot 2.776\right] \approx [1.782, 2.258]}$$

   as a 95%-confidence interval for μ.

(b) Explain why for general $X_1, \dots, X_n \overset{\text{i.i.d}}{\sim} \mathcal{N}(\mu, \sigma^2)$ random variables, $\alpha \in (0,1)$, the interval

$$S(\mathbf{X}) = \left[\frac{n-1}{\chi^2_{n-1,1-\frac{\alpha}{2}}} S_n^2, \frac{n-1}{\chi^2_{n-1,\frac{\alpha}{2}}} S_n^2\right], \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

is a $(1-\alpha)$-confidence interval for σ^2, when $\chi^2_{k,\beta}$ describes the β-quantile of the χ^2 distribution with k degrees of freedom. Calculate a 95%-confidence interval for σ^2 based on the observed data in the concrete example.

Hint: Use Theorem 2.7, (i).

Note that by Theorem 2.7, (i)[1], we have

$$\mathbb{P}\left[\chi^2_{\alpha/2} \leq \frac{(n-1)S_n^2}{\sigma^2} \leq \chi^2_{1-\alpha/2}\right] = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$$

where

$$\chi^2_{\alpha/2} \leq \frac{(n-1)S_n^2}{\sigma^2} \leq \chi^2_{1-\alpha/2} \qquad \iff \frac{n-1}{\chi^2_{n-1,1-\frac{\alpha}{2}}} S_n^2 \leq \sigma^2 \leq \frac{n-1}{\chi^2_{n-1,\frac{\alpha}{2}}} S_n^2$$

thus leading us to the desired result.

From the given data points, when $\alpha = 0.05$, we have $S_n^2 = 0.037$, $\chi^2_{4,0.975} \approx 11.143$ and $\chi^2_{4,0.025} \approx 0.484$, so that

$$\boxed{S(\mathbf{X}) \approx \left[\frac{4}{11.143} \cdot 0.037, \frac{4}{0.484} \cdot 0.037\right] \approx [0.013, 0.306]}$$

[1]Let $X_1, \dots, X_n \overset{\text{i.i.d}}{\sim} \mathcal{N}(\mu, \sigma^2)$ random variables. Then $(n-1)S_n^2/\sigma^2 \sim \chi^2_{n-1}$.

2. **Asymptotic confidence interval for the Poisson distribution**

Let $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} Pois(\lambda)$, with unknown $\lambda > 0$.

(a) Calculate the Fisher information $I(\lambda)$.

For Poisson distribution, $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$, and we have $\mathbb{E}[X] = \lambda$, $\mathbb{E}[X^2] = \lambda^2 + \lambda$. Hence, by definition, we have

$$\begin{aligned}
I(\lambda) &= \mathbb{E}_\lambda\left[\left(\frac{d \log f_\lambda(X_1)}{d\lambda}\right)^2\right] \\
&= \mathbb{E}_\lambda\left[\left(\frac{d}{d\lambda}\left(X_1 \log(\lambda) - \lambda - \log(X_1!)\right)\right)^2\right] \\
&= \mathbb{E}_\lambda\left[\left(\frac{X_1}{\lambda} - 1\right)^2\right] \\
&= \mathbb{E}_\lambda\left[\frac{X_1^2}{\lambda^2} - \frac{2X_1}{\lambda} + 1\right] \\
&= \mathbb{E}_\lambda\left[\frac{X_1^2}{\lambda^2}\right] - \mathbb{E}_\lambda\left[\frac{2X_1}{\lambda}\right] + 1 \\
&= 1 + \frac{1}{\lambda} - 2 + 1 = \boxed{\frac{1}{\lambda}}
\end{aligned}$$

(b) For $\alpha \in (0, 1)$, find an asymptotic $(1-\alpha)$-confidence interval for λ.

By Theorem 6.7, HW 5.1 (b)[2], and (a), we have

$$\begin{aligned}
S(\mathbf{X}) &= \left[\widehat{\lambda}_n - \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{nI\left(\widehat{\lambda}_n\right)}}, \widehat{\lambda}_n + \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{nI\left(\widehat{\lambda}_n\right)}}\right] \\
&= \left[\widehat{\lambda}_n - \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{\frac{n}{\widehat{\lambda}_n}}}, \widehat{\lambda}_n + \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{\frac{n}{\widehat{\lambda}_n}}}\right] \\
&= \boxed{\left[\widehat{\lambda}_n - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{\lambda}_n}{n}}, \widehat{\lambda}_n + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{\lambda}_n}{n}}\right], \qquad \widehat{\lambda}_n = \overline{X}_n,}
\end{aligned}$$

where $u_{1-\gamma}$ denotes the $(1-\gamma)$-quantile of the $\mathcal{N}(0,1)$-distribution, as an asymptotic $(1-\alpha)$-confidence interval for λ.

[2] MLE of λ is $\widehat{\lambda}_n = \overline{X}_n$.

3. **Asymptotic confidence interval for the Exponential distribution**

Let $X_1, \ldots, X_n \stackrel{\text{i.i.d}}{\sim} \mathcal{E}(\theta)$, with unknown $\theta > 0$.

(a) Calculate the Fisher information $I(\theta)$.

For exponential distribution, $f(x) = \theta e^{-\theta x}, x \geq 0$, and we have $\mathbb{E}[X] = 1/\theta$, $\mathbb{E}\left[X^2\right] = 2/\theta^2$. Hence, by definition, we have

$$\begin{aligned} I(\theta) &= \mathbb{E}_\theta\left[\left(\frac{d \log f_\theta(X_1)}{d\theta}\right)^2\right] \\ &= \mathbb{E}_\theta\left[\left(\frac{d}{d\theta}(\log\theta - \theta X_1)\right)^2\right] \\ &= \mathbb{E}_\theta\left[\left(\frac{1}{\theta} - X_1\right)^2\right] \\ &= \mathbb{E}_\theta\left[\frac{1}{\theta^2} - \frac{2X_1}{\theta} + X_1^2\right] \\ &= \mathbb{E}_\theta\left[X_1^2\right] - \mathbb{E}_\theta\left[\frac{2X_1}{\theta}\right] + \frac{1}{\theta^2} \\ &= \frac{2}{\theta^2} - \frac{2}{\theta^2} + \frac{1}{\theta^2} = \boxed{\frac{1}{\theta^2}} \end{aligned}$$

(b) For $\alpha \in (0,1)$, find an asymptotic $(1-\alpha)$-confidence interval for θ.

By Theorem 6.7, Example 4.4 (ii)[3], and (a), we have

$$\begin{aligned} S(\mathbf{X}) &= \left[\widehat{\theta}_n - \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{nI\left(\widehat{\theta}_n\right)}}, \widehat{\theta}_n + \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{nI\left(\widehat{\theta}_n\right)}}\right] \\ &= \left[\widehat{\theta}_n - \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{\frac{n}{\widehat{\theta}_n^2}}}, \widehat{\theta}_n + \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{\frac{n}{\widehat{\theta}_n^2}}}\right] \\ &= \boxed{\left[\widehat{\theta}_n - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{\theta}_n^2}{n}}, \widehat{\theta}_n + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{\theta}_n^2}{n}}\right], \qquad \widehat{\theta}_n = \frac{1}{\overline{X}_n},} \end{aligned}$$

where $u_{1-\gamma}$ denotes the $(1-\gamma)$-quantile of the $\mathcal{N}(0,1)$-distribution, as an asymptotic $(1-\alpha)$-confidence interval for θ.

[3] MLE of θ is $\widehat{\theta}_n = \frac{1}{\overline{X}_n}$.

4. **Exact confidence interval for the Exponential distribution**

Let $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \mathcal{E}(\theta)$, with unknown $\theta > 0$.

(a) Show that $2\theta \sum_{i=1}^{n} X_i$ follows a χ^2_{2n} distribution.

We show that the sum of n i.i.d *exponential* random variables follows the Erlang (θ, n) distribution[4], which is then a χ^2_{2n} distribution after multiplying 2θ.

Theorem. *Let $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathcal{E}(\theta)$. $\sum_{i=1}^{n} X_i$ follows an Erlang (θ, n) distribution.*

Proof. By definition, for $i = 1, \ldots, n$, we have $f_{X_i}(x) = \theta e^{-\theta x}, x \geq 0$. One can show[5] that its Moment generating function is $M_{X_i}(t) = \frac{\theta}{\theta - t}, t < \theta$. Let $X = \sum_{i=1}^{n} X_i$. We have

$$
\begin{aligned}
M_X(t) &= \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{t\sum_{i=1}^{n} X_i}\right] \\
&= \mathbb{E}\left[e^{tX_1}\right]\mathbb{E}\left[e^{tX_2}\right]\ldots\mathbb{E}\left[e^{tX_n}\right] \\
&= M_{X_1}(t)M_{X_2}(t)\ldots M_{X_n}(t) \\
&= \left(\frac{\theta}{\theta - t}\right)^n, \qquad t < \theta,
\end{aligned}
$$

which is the moment generation function of an Erlang (θ, n) random variable. □

Theorem. *If $X \sim$ Erlang (θ, n), then $2\theta X \sim \chi^2_{2n}$.*

Proof. The transformation $Y = g(X) = 2\theta X$ is a 1-1 transformation from $\{x | x > 0\}$ to $\{y | y > 0\}$ with inverse $X = g^{-1}(Y) = Y/(2\theta)$ and Jacobian $dX/dY = 1/(2\theta)$. Hence, by the transformation technique, the probability density function of Y is

$$
\begin{aligned}
f_Y(y) &= f_X\left(g^{-1}(y)\right)|dx/dy| \\
&= \theta^n \frac{(y/(2\theta))^{n-1} e^{-\theta y/(2\theta)}}{(n-1)!}\frac{1}{2\theta} \qquad (\theta > 0) \\
&= \frac{y^{n-1} e^{-y/2}}{2^n (n-1)!} = \frac{y^{n-1} e^{-y/2}}{2^n \Gamma(n)}, \qquad y > 0,
\end{aligned}
$$

which is the probability density function of the χ^2_{2n} distribution. □

[4] An Erlang (θ, n) is a special case of *Gamma* distribution. It has probability density function

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

[5] Moment Generating Function of Exponential Distribution.

(b) Use (a) to find an *exact* $(1-\alpha)$-confidence interval for θ, for $\alpha \in (0,1)$.

Note that

$$\mathbb{P}\left[\chi^2_{2n,\alpha/2} \le 2\theta \sum_{i=1}^n X_i \le \chi^2_{2n,1-\alpha/2}\right] = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$$

where

$$\chi^2_{2n,\alpha/2} \le 2\theta \sum_{i=1}^n X_i \le \chi^2_{2n,1-\alpha/2} \qquad \Longleftrightarrow \frac{\chi^2_{2n,\alpha/2}}{2\sum_{i=1}^n X_i} \le \theta \le \frac{\chi^2_{2n,1-\alpha/2}}{2\sum_{i=1}^n X_i}$$

Hence

$$\boxed{S(\mathbf{X}) = \left[\frac{\chi^2_{2n,\alpha/2}}{2\sum_{i=1}^n X_i}, \frac{\chi^2_{2n,1-\alpha/2}}{2\sum_{i=1}^n X_i}\right]}$$

is a $(1-\alpha)$-confidence interval for θ, for $\alpha \in (0,1)$.

Problem Set 8

Submission:
Thursday, 04/28/2022, until 1 PM, to be uploaded on the NYU Brightspace course homepage.

---

1. **Testing binomial distributions I** [4 points]

   Imagine we want to test whether a given coin is fair.

   (a) Suppose that there is *no* additional information. What would be a reasonable test for at level $\alpha = 0.05$ to determine whether the coin is fair, when we flip it $n = 200$ times? Formulate H_0 and H_1 and determine the critical region.

   (b) Suppose now that we suspect that the coin has a higher chance to land on heads. How would the test in (a) change?

   (c) Consider both situations in (a) and (b), but use an approximation of the binomial distribution coming from the central limit theorem. How do the critical values change?

2. **Testing binomial distributions II** [4 Points]

   The first digit of various numerical data sets (such as electricity bills, street addresses, stock prices,...) typically follows the *Benford law*. This is a distribution on $(\{1,\dots,9\}, \mathscr{P}(\{1,\dots,9\})$ with probability mass function

   $$p(k) = \log_{10}\left(1 + \frac{1}{k}\right), \qquad 1 \leq k \leq 9.$$

   Here $\log_{10} = \frac{\log}{\log(10)}$ is the logarithm with base 10.

   (a) Show that $(p(k))_{k\in\{1,\dots,9\}}$ defines a probability mass function on $\{1,\dots,9\}$.

   By Benford's law, 1 should be the first digit in roughly 30% of the numbers in a valid statistical data set. Suppose we suspect a given sample of 100 independent data points to be fraudulent, since the first digit 1 shows up only 17 times in the sample.

   (b) Formulate a testing problem and construct the Neyman-Pearson test ϕ^* at level $\alpha = 0.05$ for the hypotheses

   H_0 : the probability for 1 as first digit is 30%, against
   H_1 : the probability for 1 as first digit is smaller than 30%.

   Find the critical region and determine whether H_0 can be rejected with our observation.

   (c) Determine the p-value of the test ϕ^* from the previous subexercise. Explain its interpretation in the context of the problem.

   (d) Suppose that it is revealed later that the data set was in fact fraudulent, and generated by some method in which the probability of having first digit 1 was only 20%. Calculate the probability of a type II error for the test ϕ^*.

3. **Neyman-Pearson test for Poisson distributions** [4 Points]

The number of claims reported to an insurance company during a year is Poisson-distributed with some parameter $\lambda > 0$. From previous years, the insurance company uses $\lambda = \lambda_0$ as a model parameter. The company notices that the number of claims has increased, and therefore wants to test

$$\begin{aligned} H_0 : &\qquad \lambda = \lambda_0, \text{ against} \\ H_1 : &\qquad \lambda = \lambda_1, \end{aligned}$$

where $\lambda_1 > \lambda_0$.

(a) Find a Neyman-Pearson test ϕ^* at level $\alpha \in (0,1)$ for this testing problem.

(b) Is the test from the previous subexercise a uniformly most powerful test for H_0 against

$$\widetilde{H}_1 : \qquad \lambda \in \{\lambda' \,;\, \lambda' > \lambda_0\}?$$

(c) Suppose now that $\lambda_0 = 9000$ and the company observed 9876 claims. Decide whether H_0 is rejected at a level $\alpha = 0.05$.

Hint: Use that $\sum_{k=0}^{9155} \frac{(9000)^k}{k!} e^{-9000} \approx 0.9491$ and $\sum_{k=0}^{9156} \frac{(9000)^k}{k!} e^{-9000} \approx 0.9502$.

4. **Testing for the variance in normal distributions** [4 Points]

Suppose that $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ with *known* $\mu \in \mathbb{R}$ and unknown $\sigma^2 > 0$. Assume furthermore that $0 < \sigma_0^2 < \sigma_1^2$. Show that the test

$$\phi^*(\mathbf{X}) = \begin{cases} 1, & \text{if } \sum_{i=1}^n (X_i - \mu)^2 > \sigma_0^2 \chi^2_{n,1-\alpha}, \\ 0, & \text{if } \sum_{i=1}^n (X_i - \mu)^2 \leq \sigma_0^2 \chi^2_{n,1-\alpha}, \end{cases}$$

is the Neyman-Pearson test at level $\alpha \in (0,1)$ for $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 = \sigma_1^2$.

Homework 8, MATH.UA 234

Rex Liu

April 27, 2022

1. **Testing binomial distributions I**

   (a) Suppose $X_i = 1$ means the coin gets head and $X_i = 0$ tail. Let $X = \sum_{i=1}^{200} X_i$. Hypothesis:

   $$\begin{aligned} H_0 &: \quad \text{the coin is fair,} \\ H_1 &: \quad \text{the coin is biased.} \end{aligned}$$

   Note that a *critical region* is a set of values for the test statistic for which the null hypothesis is rejected. We are asked to perform the test at a 5% significance level. Note that "biased" is a two-tailed test. From the cumulative binomial distribution table, we see that for $Bin(200, \frac{1}{2})$, $\mathbb{P}[X \leq 86]+\mathbb{P}[X \geq 114] = 0.056$ and $\mathbb{P}[X \leq 85]+\mathbb{P}[X \geq 115] = 0.040$. Hence,

   $$C_\alpha = \left\{ x \in \mathcal{X} : \sum_{i=1}^{n} x_i \in \{1, 2, \dots, 85, 115, 116, \dots, 200\} \right\}$$

   is the critical region for the test.

   (b) Since we suspect that the coin has a higher chance to land on heads, we change the hypothesis:

   $$\begin{aligned} H_0 &: \quad \text{the coin is fair,} \\ H_1 &: \quad \text{the coin is biased in favor of heads.} \end{aligned}$$

   It is a one-tailed test. For $Bin(200, \frac{1}{2})$, $\mathbb{P}[X \geq 112] = 0.052$ and $\mathbb{P}[X \geq 113] = 0.038$. Hence,

   $$C_\alpha = \left\{ x \in \mathcal{X} : \sum_{i=1}^{n} x_i \in \{113, 114, \dots, 200\} \right\}$$

   is the critical region for the test.

   (c) Note that for $Bin(200, \frac{1}{2})$, $\mu = 100$ and $\sigma = \sqrt{50}$. By the central limit theorem, we have $\frac{X-100}{\sqrt{50}} = \frac{X-100}{\sqrt{50}} = Y \sim \mathcal{N}(0, 1)$.

   For situation in (a), we have, from the cumulative normal distribution table, $\mathbb{P}[Y \in (-1.96, 1.96)] = 95\%$ and thus, $86.15 < X < 113.85$. The critical value is then 86 and 114, which is closer to 100 compared with the exact binomial test $(85, 115)$.

   For situation in (b), we have $\mathbb{P}[Y > 1.645] = 0.05$ and thus, $X > 111.63$. The critical value is then 112, which is also closer to 100 compared with the exact binomial test (113).

2. **Testing binomial distributions II**

(a) First, we check non-negativity. Since $(1 + 1/k) > 1$ for $k \in [1, 9]$, $p(k)$ is non-negative by the property of logarithm. Then we show that $\sum_{k\in\{1,\ldots,9\}} p(k) = 1$.

$$\sum_{k\in\{1,\ldots,9\}} p(k) = \sum_{k\in\{1,\ldots,9\}} \log_{10}(1 + 1/k)$$
$$= \log_{10}\left(\prod_{k\in\{1,\ldots,9\}} (1 + 1/k)\right) = \log_{10}\left(2 \cdot \frac{3}{2} \cdot \ldots \cdot \frac{10}{9}\right) = 1$$

Hence, the PMF is well-defined.

(b) Formulate: we consider 1 as the first digit in roughly 30% of the numbers in a valid statistical data set. We want to see whether the probability is less than 30% for a potentially fraudulent sample of 100 independent data points. We choose the model $X_1, \ldots, X_n \sim Ber(\theta)$. $X_i = 1$ if the first digit is 1 while $X_i = 0$ if the first digit is not 1. Question: do we have

$$\begin{aligned} &\mathbb{P}_{\theta_0} \quad (\mathbb{P}_{\theta_0})_{X_1} = Ber(\theta_0), \quad \theta_0 = 0.3 && (H_0) \\ \text{or } &\mathbb{P}_{\theta_1} \quad (\mathbb{P}_{\theta_1})_{X_1} = Ber(\theta_1), \quad \theta_1 \in [0, 0.3) && (H_1) \end{aligned}$$

Construct: similar to *Example 7.8* in notes, we have

$$L(\mathbf{x}) = \left(\frac{\theta_1}{1-\theta_1} / \frac{\theta_0}{1-\theta_0}\right)^{\sum_{i=1}^n x_i} \left(\frac{1-\theta_1}{1-\theta_0}\right)^n \geq c$$

Since $\theta_1 < \theta_0$, we have $\frac{\theta_1}{1-\theta_1} / \frac{\theta_0}{1-\theta_0} < 1$, so

$$L(\mathbf{x}) \geq c \quad \Longleftrightarrow \quad \sum_{i=1}^n x_i \leq k$$

We choose k for $n = 100$, $\theta_0 = 0.3$ and $\alpha \leq 0.05$ such that $\mathbb{P}_{\theta_0}\left[\sum_{i=1}^n X_i \leq k\right] = \alpha$. Note that sum of 100 $X_i \sim Ber(0.3)$ follows $Bin(100, 0.3)$ distribution. We see that $\mathbb{P} = 0.048$ for $k = 22$ and $\mathbb{P} = 0.076$ for $k = 23$. Hence, we choose $k = 22$, and the Neyman-Pearson test $\phi^* \circ \mathbf{X}$ is given by

$$\phi^*(\mathbf{x}) = \begin{cases} 1, & \sum_{i=1}^n x_i \leq 22 \\ 0, & \sum_{i=1}^n x_i > 22 \end{cases}$$

The critical region is

$$C_\alpha = \left\{x \in \mathcal{X} : \sum_{i=1}^n x_i \in \{0, 1, \ldots, 22\}\right\}$$

Since 17 is in the critical region, H_0 can be rejected.

(c) Note that we need 17 in the critical region to reject H_0, and note that $\mathbb{P} = 0.00216$ for $k = 17$, so that p-value $= 0.00216$.

Interpretation: $\alpha = 0.00216$ is the smallest value such that we can reject H_0 based on the observation that *the first digit* 1 *shows up only* 17 *times in the sample.*

(d) Given $\theta_1 = 0.2$, we have the type II error

$$\mathbb{P}_{\theta_1}[\phi * (\mathbf{X}) = 0] = \mathbb{P}_{\theta_1}\left[\sum_{i=1}^{n} X_i \geq 23\right] = 0.26107.$$

3. **Neyman-Pearson test for Poisson distributions**

(a) Let X be the number of claims. The likelihood quotient is given as

$$\begin{aligned} L(x) &= \frac{f_{\lambda_1}(x)}{f_{\lambda_0}(x)} \\ &= \frac{\mathbb{P}[X \approx x|\lambda = \lambda_1]}{\mathbb{P}[X \approx x|\lambda = \lambda_0]} \\ &= \frac{\lambda_1^x e^{-\lambda_1}}{x!} \frac{x!}{\lambda_0^x e^{-\lambda_0}} \\ &= \left(\frac{\lambda_1}{\lambda_0}\right)^x e^{\lambda_0 - \lambda_1} \end{aligned}$$

Hence, the Neyman-Pearson test $\phi^* \circ \mathbf{X}$ is given by

$$\phi^*(x) = \begin{cases} 1, & L(x) \geq c \\ 0, & L(x) < c \end{cases}$$

where c is a constant chosen by the significance level

$$\alpha = \mathbb{P}_{\lambda_0}\left[\left(\frac{\lambda_1}{\lambda_0}\right)^x e^{\lambda_0 - \lambda_1} \geq c\right]$$

(b) *It is.* Since the likelihood ratio $L(x)$ is monotone in x, it is valid to deduce that (also note that $\lambda_1 > \lambda_0$ by assumption)

$$\left(\frac{\lambda_1}{\lambda_0}\right)^x e^{(\lambda_0 - \lambda_1)} \geq c \quad \iff \quad x \geq k$$

for some well-chosen k. Hence the only information we used here is that $\lambda_1 > \lambda_0$, so it is a UMP test for H_0 against $\tilde{H}_1$.

(c) Given $\lambda_0 = 9000, \alpha = 0.05$, we calculate k by $\mathbb{P}_{\lambda_0}[x \geq k] = \alpha$. By the hint, we have $\mathbb{P}_{\lambda_0}[x \leq 9155] = 0.9491$ while $\mathbb{P}_{\lambda_0}[x \leq 9156] = 0.9502$. Since we seek for the largest $\alpha \leq 0.05$, we let $k = 9157$. Note that $9876 > 9157$, hence H_0 is rejected.

4. **Testing for the variance in normal distributions**

Proof. Note that

$$\begin{aligned} L(x_1, \dots, x_n) &= \frac{\mathbb{P}[X_1 \approx x_1, \dots, X_n \approx x_n | \sigma = \sigma_1]}{\mathbb{P}[X_1 \approx x_1, \dots, X_n \approx x_n | \sigma = \sigma_0]} \\ &= \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x_i-\mu)^2/2\sigma_1^2}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_0} e^{-(x_i-\mu)^2/2\sigma_0^2}} \\ &= \left(\frac{\sigma_0}{\sigma_1}\right)^n \exp\left\{\frac{\sigma_1^2 - \sigma_0^2}{2\sigma_0^2\sigma_1^2} \sum_{i=1}^n (x_i - \mu)^2\right\} \end{aligned}$$

Since $\sigma_0^2 < \sigma_1^2$, we have $L(\mathbf{x}) \geq c \iff \sum_{i=1}^n (x_i - \mu)^2 \geq \tilde{c}$ where $\tilde{c} = \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2}\left(n \ln \frac{\sigma_1}{\sigma_0} + \ln c\right)$. Under the hypothesis $\sigma = \sigma_0$, we have $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma_0}\right)^2 \sim \chi_n^2$. Hence,

$$\begin{aligned} \alpha &= \mathbb{P}[L(X_1, \dots, X_n) \geq c | \sigma = \sigma_0] \\ &= \mathbb{P}\left[\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma_0}\right)^2 \geq \frac{\tilde{c}}{\sigma_0^2} \middle| \sigma = \sigma_0\right] \end{aligned}$$

which gives us $\tilde{c} = \sigma_0^2 \chi_{n,1-\alpha}^2$. Hence $\phi^*(\mathbf{X}) = 1$ when $\sum_{i=1}^n (X_i - \mu)^2 \geq \sigma_0^2 \chi_{n,1-\alpha}^2$ and $\phi^*(\mathbf{X}) = 0$ otherwise.

Note that $>$ and $\geq$ here are trivial as the chi-square distribution is continuous, but following the class notes, we should use $\geq$ instead of $>$ in the problem. □

MATH-UA.0234 | Mathematical Statistics | Spring 2022
Dr. M. Nitzschner

Solutions to Problem Set 9

---

1. **Z- and t-test** [4 Points]

Suppose it is known that the mean temperature during a given year in a city at the beginning of the 19th century was $52.2°$ F. We suspect that the mean temperature in this city has increased since then. Suppose we measure the mean temperatures in the years 2010 and 2018 (in $°$ F) in this city and find:

$$54.0 \quad 52.3 \quad 53.1 \quad 53.2 \quad 52.0 \quad 52.2 \quad 52.9 \quad 53.4 \quad 52.7.$$

Assume that these temperatures can be viewed approximately as realizations of i.i.d. normally distributed random variables, i.e. $X_1, ..., X_9 \sim \mathcal{N}(\mu, \sigma^2)$.

(a) Construct a test at level $\alpha = 0.05$ for the hypothesis

$$H_0 : \mu = \mu_0 = 52.2 \qquad \text{against} \qquad H_1 : \mu > \mu_0 = 52.2,$$

if the standard deviation $\sigma = 1$ is *known*. Determine the p-value of the test and whether H_0 is rejected at level $\alpha = 0.05$.

(b) Construct a test at level $\alpha = 0.05$ for the same hypothesis as above, but in the situation where σ is *unknown*. Determine the p-value of the test and whether H_0 is rejected at level $\alpha = 0.05$.

(a) We have seen for the given situation, the one-sided Z-test (Theorem 7.12) can be applied, and it is given by

$$\phi^*(X_1, ..., X_n) = \begin{cases} 1, & \overline{X}_n \geq k^*_\alpha, \\ 0, & \overline{X}_n < k^*_\alpha, \end{cases}$$

where

$$k^*_\alpha = \mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}.$$

We can calculate

$$\overline{X}_9 = 52.87, \qquad S_9^2 = 0.41.$$

We then see that

$$k^*_{0.05} = \mu_0 + \frac{\sigma}{\sqrt{n}} u_{0.95} = 52.2 + \frac{1}{\sqrt{9}} \cdot 1.645 \approx 52.75.$$

Since $\overline{X}_9 > k^*_{0.05}$, we see that the null hypothesis is rejected. For the p-value, we need to calculate the probability to obtain a value as least as extreme as the observed one, so

$$\begin{aligned} p\text{-value} &= \mathbf{P}_{\mu_0}\left[\overline{X}_n \geq 52.87\right] \\ &= \mathbf{P}_{\mu_0}\left[\sqrt{n}\frac{\overline{X}_n - 50.7}{1} \geq \sqrt{n}\frac{52.87 - 52.2}{1}\right] = 1 - \Phi(2.01) \approx 0.022. \end{aligned}$$

(b) Now we need to use the (one-sided) t-test (Theorem 7.14), which is given by

$$\phi^*(X_1, ..., X_n) = \begin{cases} 1, & \frac{\overline{X}_n - \mu_0}{S_n/\sqrt{n}} \geq k^*_\alpha, \\ 0, & \frac{\overline{X}_n - \mu_0}{S_n/\sqrt{n}} < k^*_\alpha, \end{cases}$$

where k^*_α is the $(1-\alpha)$-quantile of the t_{n-1}-distribution. In our case, we have

$$k^*_\alpha = t_{9-1,0.95} = 1.86,$$

and we calculate

$$\frac{\overline{X}_n - \mu_0}{S_n/\sqrt{n}} = \frac{52.87 - 52.2}{\sqrt{0.41}/\sqrt{9}} \approx 3.141 > 1.86.$$

And H_0 is again rejected. For the p-value, we need to calculate the probability to obtain a value at least as extreme as the observed one, namely

$$p\text{-value} = \mathbf{P}_{\mu_0}\left[\frac{\overline{X}_9 - \mu_0}{S_9/\sqrt{9}} \geq 3.141\right] = 1 - \mathbf{P}[T \leq 3.141] \approx 0.0069, \qquad T \sim t_8,$$

2. **Pearson's χ^2-test** [4 points]

A small company wants to test at level $\alpha = 0.05$ whether the number of work orders is equally distributed among the five workdays of the week. The number of work orders is recorded over a certain time period and the following is found:

Monday	Tuesday	Wednesday	Thursday	Friday
97	71	60	64	82

(a) Formulate H_0 and H_1 for this problem. Using Pearson's χ^2-test, can H_0 be rejected at level $\alpha = 0.05$?

(b) Calculate the p-value for the given observation.

(a) We have the null and alternative hypotheses

$$\begin{aligned} H_0: &\qquad p_j = \pi_j = \frac{1}{5}, &&\qquad 1 \leq j \leq 5, \\ H_1: &\qquad p_j \neq \pi_j = \frac{1}{5}, &&\qquad \text{for some } j \in \{1,2,3,4,5\}. \end{aligned}$$

We perform a χ^2-test at level $\alpha = 0.05$ (Theorem 7.23), which is given by

$$\phi(\mathbf{X}) = \begin{cases} 1, & \sum_{i=1}^r \frac{(N_i - n\pi_i)^2}{n\pi_i} \geq \chi^2_{r-1,1-\alpha}, \\ 0, & \sum_{i=1}^r \frac{(N_i - n\pi_i)^2}{n\pi_i} < \chi^2_{r-1,1-\alpha}. \end{cases} \tag{1}$$

In our situation we have $r = 5$ and $\alpha = 0.05$, so we need

$$\chi^2_{4,0.95} = 9.488.$$

Note that $n = 97 + 71 + 60 + 64 + 82 = 374$. Let us calculate the relevant test statistic

$$\begin{aligned} \sum_{i=1}^r \frac{(N_i - n\pi_i)^2}{n\pi_i} &= \frac{(97 - 374\cdot\frac{1}{5})^2}{374\cdot\frac{1}{5}} + \frac{(71 - 374\cdot\frac{1}{5})^2}{374\cdot\frac{1}{5}} + \frac{(60 - 374\cdot\frac{1}{5})^2}{374\cdot\frac{1}{5}} \\ &\quad + \frac{(64 - 374\cdot\frac{1}{5})^2}{374\cdot\frac{1}{5}} + \frac{(82 - 374\cdot\frac{1}{5})^2}{374\cdot\frac{1}{5}} \approx 11.96 > 9.488 = \chi^2_{4,0.95}, \end{aligned}$$

so we can reject H_0. For the p-value, we again calculate the probability that the test statistic attains a value at least as extreme as 11.96, so

$$p\text{-value} = \mathbf{P}[C \geq 11.96] \approx 0.018, \qquad C \sim \chi^2_{4,0.95}.$$

3. **Bayesian inference I** [4 Points]

Let $X \sim Bin(N,p)$ with $p \in (0,1)$ *known* and N a prior with Poisson distribution $N \sim Pois(\lambda)$, $\lambda > 0$.

(a) Calculate the posterior distribution of N.

(b) Determine the corresponding posterior expectation.

(a) We use the shorthand notation from the lecture (Remark 8.2 (ii)) for prior, likelihood and posterior:

$$[n] = \frac{\lambda^n}{n!}e^{-\lambda} \propto \frac{\lambda^n}{n!},$$
$$[x|n] = \binom{n}{x}p^x(1-p)^{n-x} \propto \frac{n!}{(n-x)!}(1-p)^{n-x}, \qquad n \geq x:$$

Therefore:

$$[n|x] \propto [n]\cdot[x|n] \propto \frac{\lambda^n}{n!}\cdot\frac{n!}{(n-x)!}(1-p)^{n-x} = \frac{\lambda^n(1-p)^{n-x}}{(n-x)!} \propto \frac{(\lambda(1-p))^{n-x}}{(n-x)!}, \qquad n \geq x.$$

This itself is proportional to a Poisson distribution is with parameter $\lambda(1-p)$, shifted by x, so

$$N|X = x \sim Pois(\lambda(1-p)) + x.$$

Comment: For completeness, let us also show the complete calculation without arguments of proportionality.

$$[n|x] = \mathbf{P}[N = n|X = x] = \frac{[n]\cdot[x|n]}{\sum_{k=0}^{\infty}[k]\cdot[x|k]} = \frac{\mathbf{P}[N = n]\cdot\mathbf{P}[X = x|N = n]}{\sum_{k=0}^{\infty}\mathbf{P}[N = k]\cdot\mathbf{P}[X = x|N = k]}$$
$$= \frac{\frac{\lambda^n}{n!}e^{-\lambda}\cdot\binom{n}{x}p^x(1-p)^{n-x}}{\sum_{k=x}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}\cdot\binom{k}{x}p^x(1-p)^{k-x}}, \qquad n \geq x$$
$$= \frac{\lambda^n\cdot\frac{1}{x!(n-x)!}(1-p)^{n-x}}{\sum_{k=x}^{\infty}\lambda^k\cdot\frac{1}{x!(k-x)!}(1-p)^{k-x}}, \qquad n \geq x$$
$$= \frac{\lambda^{n-x}}{(n-x)!}\cdot(1-p)^{n-x}\frac{1}{\sum_{k=x}^{\infty}\frac{(\lambda(1-p))^{k-x}}{(k-x)!}}, \qquad n \geq x$$
$$= \frac{(\lambda(1-p))^{n-x}}{(n-x)!}e^{-\lambda(1-p)}, \qquad n \geq x.$$

(b) We have

$$\mathbf{E}[N|X = x] = x + \lambda(1-p).$$

4. **Bayesian inference II** [4 Points]

Recall that the Γ-distribution with parameters $\alpha, \beta > 0$ is a continuous probability distribution with probability density function

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}\mathbb{1}_{[0,\infty)}(x).$$

Suppose that $X_1, ..., X_n|\lambda$ are i.i.d. with $X_1 \sim Pois(\lambda)$.

(a) Let $\lambda \in \Gamma(\alpha, \beta)$. Calculate the posterior distribution of λ given $X_1 = x_1, ..., X_n = x_n$. Which property does the class of Γ distributions have with respect to the family of likelihoods $\{[x_1, ..., x_n|\lambda] : \lambda > 0\}$?

(b) Calculate the Jeffreys-prior for λ and determine the corresponding posterior distribution for λ given $X_1 = x_1, ..., X_n = x_n$.

(a) As in Problem 3, we use the shorthand notation for prior, likelihood and posterior. We have

$$[\lambda] = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}\mathbb{1}_{[0,\infty)}(\lambda) \propto \lambda^{\alpha-1}e^{-\beta\lambda}\mathbb{1}_{[0,\infty)}(\lambda),$$
$$[x_1, ..., x_n|\lambda] = \prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda} \propto e^{-n\lambda}\lambda^{\sum_{i=1}^{n}x_i},$$

We obtain

$$[\lambda|x_1,...,x_n] \propto [\lambda]\cdot[x_1,...,x_n|\lambda]$$
$$\propto \lambda^{\alpha-1}e^{-\beta\lambda}\mathbb{1}_{[0,\infty)}(\lambda)\cdot e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i}$$
$$\propto \lambda^{\alpha+\sum_{i=1}^n x_i-1}e^{-(n+\beta)\lambda}.$$

In other words, we have

$$\lambda \,|\, X_1 = x_1,...,X_n = x_n \sim \Gamma\left(\alpha+\sum_{i=1}^n x_i, n+\beta\right).$$

This means that the class of Γ-distributions is conjugate with respect to the family of Poisson likelihoods.

(b) The Jeffreys-prior is calculated by using

$$[\lambda] \propto \sqrt{I(\lambda)},$$

where $I(\lambda)$ is the Fisher information. We use

$$I(\lambda) = \mathbf{E}_\lambda\left[-\frac{\partial^2}{\partial\lambda^2}\log p_\lambda(X)\right] \overset{\text{Problem set 7, P.2}}{=} \frac{1}{\lambda}.$$

In other words, the Jeffreys prior is

$$[\lambda] \propto \frac{1}{\sqrt{\lambda}}$$

(note that this is an improper prior). Now for the posterior,

$$[\lambda|x_1,...,x_n] \propto [\lambda]\cdot[x_1,...,x_n|\lambda] \propto \lambda^{-\frac{1}{2}}\cdot e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i} = e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i+\frac{1}{2}-1}$$

so we find

$$\lambda \,|\, X_1 = x_1,...,X_n = x_n \sim \Gamma\left(\sum_{i=1}^n x_i+\frac{1}{2}, n\right).$$