DPT Exercise 9

github.com/aiulus/dpt-ss25

START: 12:15

Topics:

E1) (Properties of) Char. functions; CLT

E2) (Properties of the sample mean);
Chebyshev's inequality; Interval estimates

E3) Change of variables formula; Conv. in distr.

E4) Estimators; MLE

DPT Exercises

Exercise sheet 9 Niki Kilbertus

Exercise 1 (sum of waiting times)**.** The time (in minutes) a customer i has to wait for their coffee at a specialty coffee shop is modeled by an exponential random variable $X_i \sim \mathrm{Exp}(\lambda)$ with $\lambda = 0.5$ (so the average wait is $1/\lambda = 2$ minutes). Let $S_{36} = \sum_{i=1}^{36} X_i$ be the total time required to serve 36 *independent* customers.

(i) Find the characteristic function of a single waiting time X_i.

(ii) Using the result from (i), what is the exact distribution of the total waiting time S_{36}? State its name and parameters.
[Hint: Look at Example 2.50 of the lecture.]

(iii) Use the clt to approximate the probability that the total waiting time for 36 customers exceeds 80 minutes.
[Hint: You will need the mean and variance of an $\mathrm{Exp}(0.5)$ distribution. For the clt, you may use that for a standard normal RVRV Z, we have $P(Z > 2/3) \approx 0.2525$.]

Exercise 2 (quality control in manufacturing)**.** A factory produces resistors. The resistance of each resistor is intended to be 100 ohms. Due to manufacturing variability, the actual resistance values are iid random variables with mean $\mu = 100$ ohms and standard deviation $\sigma = 2$ ohms. A batch consists of $n = 64$ resistors. A batch is rejected if the average resistance of the 64 resistors is outside the range $[99.5, 100.5]$ ohms.

(i) Let $\bar{X}_{64}$ be the average resistance of a batch of 64 resistors. What are the mean and variance of $\bar{X}_{64}$?

(ii) Use Chebyshev's inequality to find an upper bound on the probability that a batch is rejected.

(iii) Use the clt to approximate the probability that a batch is rejected.
[Hint: You may use that for a standard normal RVRV Z, we have $P(Z > 2) \approx 0.0228$.]

Exercise 3 (transformation and convergence in distribution)**.** Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of RVRVs with $X_n \sim \mathrm{Unif}(0, 1/n)$ for each $n \in \mathbb{N}_{>0}$. Let $Y_n = nX_n$.

(i) For a fixed n, find the pdf $p_{Y_n}(y)$ of Y_n.
[Hint: Use the change of variables formula.]

(ii) Find the cdf $F_{Y_n}(y)$ of Y_n.

(iii) Determine the limiting cdf $F_Y(y) = \lim_{n\to\infty} F_{Y_n}(y)$.

(iv) To what RVRV Y does Y_n converge in distribution? Specify its distribution.

Exercise 4 (mle for uniform distribution)**.** Let $X_1, \ldots, X_n$ be iid RVRVs from a uniform distribution $\mathrm{Unif}(0, \theta)$ with unknown parameter $\theta > 0$. The pdf of X_i is $p(x_i \mid \theta) = \frac{1}{\theta}\chi_{[0,\theta]}(x_i)$.

(i) Write down the likelihood function $L(\theta \mid x)$ for an observed sample $x = (x_1, \ldots, x_n)$. Be careful with the indicator functions.

(ii) Find the maximum likelihood estimator (mle) $\hat{\theta}_{\mathrm{mle}}$ for θ.
[Hint: The likelihood function may not be differentiable with respect to θ, but the maximum value can be found by reasoning about monotonicity and where it is non-zero.]

(iii) Calculate the expected value $\mathbb{E}[\hat{\theta}_{\mathrm{mle}}]$. Is $\hat{\theta}_{\mathrm{mle}}$ an unbiased estimator for θ?
[Hint: First find the cdf of $\hat{\theta}_{\mathrm{mle}}$, then its pdf, and then compute the expectation.]

Exercise 1 (sum of waiting times). The time (in minutes) a customer i has to wait for their coffee at a specialty coffee shop is modeled by an exponential random variable $X_i \sim \mathrm{Exp}(\lambda)$ with $\lambda = 0.5$ (so the average wait is $1/\lambda = 2$ minutes). Let $S_{36} = \sum_{i=1}^{36} X_i$ be the total time required to serve 36 *independent* customers.

(i) Find the characteristic function of a single waiting time X_i.

$$\varphi_{X_i}(t) = \mathbb{E}\left[e^{itX_i}\right] = \int_{\mathbb{R}} e^{itx} \cdot \lambda \cdot e^{-\lambda x} \cdot \chi_{\mathbb{R}_{\geq 0}}(x)\, dx$$

$X_i \sim \mathrm{Exp}(\lambda)$

$$f_{\mathrm{Exp}(\lambda)}(y) = \lambda \cdot e^{-\lambda y} \cdot \chi_{\mathbb{R}_{\geq 0}}(y)$$

$$= \lambda \int_0^\infty e^{(it-\lambda)x}\, dx$$

$$= \lambda \cdot \left[\frac{1}{it-\lambda} \cdot e^{(it-\lambda)x}\right]_0^\infty = \lambda \cdot \left[0 - \frac{(-1)}{\lambda - it} \cdot e^0\right] = \frac{\lambda}{\lambda - it}\Bigg|_{\lambda = 1/2}$$

$$= \frac{1}{1-2it} = (1-2it)^{-1}$$

Exercise 1 (sum of waiting times). The time (in minutes) a customer i has to wait for their coffee at a specialty coffee shop is modeled by an exponential random variable $X_i \sim \text{Exp}(\lambda)$ with $\lambda = 0.5$ (so the average wait is $1/\lambda = 2$ minutes). Let $S_{36} = \sum_{i=1}^{36} X_i$ be the total time required to serve 36 *independent* customers.

PDF / CDF $\Leftrightarrow$ Distr. $\Leftrightarrow$ char. function

(ii) Using the result from (i), what is the exact distribution of the total waiting time S_{36}? State its name and parameters.
[Hint: Look at Example 2.50 of the lecture.]

$\phi_{\text{Gamma}(\alpha,\beta)}(t) = \left(1 - \frac{it}{\beta}\right)^{-\alpha}$

Remark 2.47:

(iv) If $X \perp\!\!\!\perp Y$, then $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.

$$\Rightarrow \phi_{S_{36}}(t) = \phi_{X_1+\dots+X_{36}}(t) = \prod_{j=1}^{36} \phi_{X_j}(t) = \prod_{j=1}^{36} (1-2it)^{-1} = (1-2it)^{-36} = \left(1 - \frac{it}{1/2}\right)^{-36}$$

$\alpha = 36$

$\beta = 1/2$

$S_{36} \sim \text{Gamma}(36, 0.5)$

Exercise 1 (sum of waiting times)**.** The time (in minutes) a customer i has to wait for their coffee at a specialty coffee shop is modeled by an exponential random variable $X_i \sim \mathrm{Exp}(\lambda)$ with $\lambda = 0.5$ (so the average wait is $1/\lambda = 2$ minutes). Let $S_{36} = \sum_{i=1}^{36} X_i$ be the total time required to serve 36 *independent* customers.

(ii) Using the result from (i), what is the exact distribution of the total waiting time S_{36}? State its name and parameters.
[Hint: Look at Example 2.50 of the lecture.]

Remark 2.47:

(iv) If $X \perp\!\!\!\perp Y$, then $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.

$$\Rightarrow \phi_{S_{36}}(t) = \phi_{X_1+\dots+X_{36}}(t) = \prod_{j=1}^{36} \phi_{X_j}(t) = \prod_{j=1}^{36} (1-2it)^{-1} = (1-2it)^{-36} = \left(\frac{1}{1-\frac{it}{1/2}}\right)^{36}$$

(iii) *Gamma distribution:* For $\alpha > 0$, $\beta > 0$, the $\Gamma(\alpha, \beta)$ distribution on $\Omega = \mathbb{R}_{>0} = (0, \infty)$ is characterized via the pdf

$$p_{\Gamma(\alpha,\beta)} : \mathbb{R} \to \mathbb{R}_{\geq 0},\ x \mapsto \chi_\Omega(x) \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\,.$$

Exercise 1 (sum of waiting times). The time (in minutes) a customer i has to wait for their coffee at a specialty coffee shop is modeled by an exponential random variable $X_i \sim \text{Exp}(\lambda)$ with $\lambda = 0.5$ (so the average wait is $1/\lambda = 2$ minutes). Let $S_{36} = \sum_{i=1}^{36} X_i$ be the total time required to serve 36 *independent* customers.

$$Z_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{var}[S_n]}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

(iii) Use the clt to approximate the probability that the total waiting time for 36 customers exceeds 80 minutes.
[Hint: You will need the mean and variance of an $\text{Exp}(0.5)$ distribution. For the clt, you may use that for a standard normal RVRV Z, we have $P(Z > 2/3) \approx 0.2525$.]

$$P(S_{36} > 80) = P\left(Z_{36} > \frac{80-72}{12}\right) = P(Z_{36} > 2/3) = 0.\overline{25}$$

$$(= 1 - \Phi(2/3))$$

$$Z_{36} = \frac{S_{36} - \mu_{S_{36}}}{\sigma_{36}} = \frac{S_{36} - 72}{12} \xrightarrow{d} \mathcal{N}(0,1)$$

$$\mathbb{E}[S_{36}] = 36 \cdot \mathbb{E}[X_i] = 72 =: \mu_{S_{36}}$$

$$\text{Var}[S_{36}] = 36 \cdot \text{Var}[X_i] = 36 \cdot 4 =: \sigma_{S_{36}}$$

Exercise 2 (quality control in manufacturing). A factory produces resistors. The resistance of each resistor is intended to be 100 ohms. Due to manufacturing variability, the actual resistance values are iid random variables with mean $\mu = 100$ ohms and standard deviation $\sigma = 2$ ohms. A batch consists of $n = 64$ resistors. A batch is rejected if the average resistance of the 64 resistors is outside the range $[99.5, 100.5]$ ohms.

(i) Let $\bar{X}_{64}$ be the average resistance of a batch of 64 resistors. What are the mean and variance of $\bar{X}_{64}$?

$$\bar{X}_{64} = \frac{1}{64} \cdot \sum_{i=1}^{64} X_i$$

$$\begin{cases} \mathbb{E}[\bar{X}_{64}] = \mathbb{E}[X_i] = 100 \quad \text{(unbiased estimate for mean)} \\ Var[\bar{X}_{64}] = Var\left[\frac{1}{64} \cdot \sum_{j=1}^{64} X_j\right] = \frac{1}{64^2} \cdot 64 \cdot Var[X_i] = \frac{2^2}{64} = 1/16 \end{cases}$$

Exercise 2 (quality control in manufacturing). A factory produces resistors. The resistance of each resistor is intended to be 100 ohms. Due to manufacturing variability, the actual resistance values are iid random variables with mean $\mu = 100$ ohms and standard deviation $\sigma = 2$ ohms. A batch consists of $n = 64$ resistors. A batch is rejected if the average resistance of the 64 resistors is outside the range $[99.5, 100.5]$ ohms.

(i) $\mathbb{E}[\bar{X}_{64}] = 100$

$\mathrm{Var}[\bar{X}_{64}] = 1/16$

(ii) Use Chebyshev's inequality to find an upper bound on the probability that a batch is rejected.

Theorem 2.40 (Chebyshev's inequality). *Let X be an RVRV with finite mean $\mathbb{E}[X]$ and finite variance $\mathrm{var}[X] > 0$. Then for any $\epsilon > 0$,*

$$P(|X - \mathbb{E}[X]| \geq \epsilon) \leq \frac{\mathrm{var}[X]}{\epsilon^2}.$$

$$P(\text{"batch rejected"}) = P(\bar{X}_{64} \notin [99.5, 100.5]) = P(\bar{X}_{64} - 100 \notin [-0.5, 0.5])$$

$$= P(|\bar{X}_{64} - \mathbb{E}[\bar{X}_{64}]| > 0.5) = \frac{\mathrm{Var}[\bar{X}_{64}]}{0.5^2} = \frac{1/16}{1/4} = 1/4 = 0.25$$

Exercise 2 (quality control in manufacturing). A factory produces resistors. The resistance of each resistor is intended to be 100 ohms. Due to manufacturing variability, the actual resistance values are iid random variables with mean $\mu = 100$ ohms and standard deviation $\sigma = 2$ ohms. A batch consists of $n = 64$ resistors. A batch is rejected if the average resistance of the 64 resistors is outside the range $[99.5, 100.5]$ ohms.

(ii) $\mathbb{E}[\bar{X}_{64}] = 100$

$Var[\bar{X}_{64}] = 1/16$

(iii) Use the clt to approximate the probability that a batch is rejected.
[Hint: You may use that for a standard normal RVRV Z, we have $P(Z > 2) \approx 0.0228$.]

$$Z_{64} = \frac{\bar{X}_{64} - \mathbb{E}[\bar{X}_{64}]}{\sqrt{Var[\bar{X}_{64}]}} = \frac{\bar{X}_{64} - 100}{1/4}$$

symmetry of $\mathcal{N}(0,1)$ $(\varphi(x))$

$$P(\bar{X}_{64} \notin [..]) = P(Z_{64} \notin [-2, 2]) = 2 \cdot P(Z_{64} > 2) = 2 \cdot (1 - \Phi(2)) \approx 0.0456$$

Exercise 3 (transformation and convergence in distribution). Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of RVRVs with $X_n \sim \mathrm{Unif}(0, 1/n)$ for each $n \in \mathbb{N}_{>0}$. Let $Y_n = nX_n$. $\Rightarrow g(X_n)$ — $g(x) = n \cdot x$

(i) For a fixed n, find the pdf $p_{Y_n}(y)$ of Y_n.
[Hint: Use the change of variables formula.]

Proposition 2.51 (transformation of densities / change of variables). *Let X be a continuous RVRV with density p_X, $g : \mathbb{R} \to \mathbb{R}$ a continuously differentiable, strictly monotonic function on the support of X, $S_X := \{x \in \mathbb{R} \mid p_X(x) > 0\}$, and set $Y := g(X)$. Then the pdf of Y is given by*

$$p_Y(y) = p_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \quad \text{for } y \in \underbrace{g(S_X)}_{\Rightarrow J_Y}.$$

For $y \notin g(S_X)$, $p_Y(y) = 0$.

„Components":

1) $p_{X_n}(x)$

2) $J_{X_n} := \{x \in \mathbb{R} \mid p_{X_n}(x) > 0\}$

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

$X \sim \mathrm{Unif}([a,b])$

PDF: $\begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$

CDF: $\begin{cases} 0 & \text{if } x < a \\ \frac{x}{b-a} & \text{if } x \in [a,b] \\ 1 & \text{if } x > b \end{cases}$

$p_{X_n}(y) = \chi_{(0,1/n)}(y) \cdot \frac{1}{1/n}$

$g(y) = n \cdot y \Rightarrow g^{-1}(y) = \frac{1}{n} \cdot y$

$\Rightarrow \frac{d}{dy} \frac{1}{n} \cdot y = 1/n$

$J_{X_n} = (0, 1/n)$

$g(X_n) = (0,1) \Rightarrow J_Y$

$$p_{Y_n}(y) = \chi_{(0,1/n)}(y/n) \cdot n \cdot \frac{1}{n} \cdot \chi_{(0,1)}(y)$$

$$= \chi_{(0,1/n)}(y/n) \cdot \chi_{(0,1)}(y)$$

$$= \mathbb{1}\{0 \le \underbrace{y/n}_{} \le 1/n\} \cdot \mathbb{1}\{0 \le y < 1\}$$

$$= \chi_{(0,1)}(y)$$

$$= \chi_{(0,1)}(y) \cdot 1 \Rightarrow Y_n \sim \mathrm{Unif}(0,1)$$

Exercise 3 (transformation and convergence in distribution). Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of RVRVs with $X_n \sim \mathrm{Unif}(0, 1/n)$ for each $n \in \mathbb{N}_{>0}$. Let $Y_n = nX_n$.

(ii) Find the cdf $F_{Y_n}(y)$ of Y_n.

$$Y_n \sim \mathrm{Unif}(0,1) \Rightarrow F_{Y_n}(y) = \begin{cases} 0 & \text{if } y < 0 \\ y & \text{if } y \in [0,1] \\ 1 & \text{if } y > 1 \end{cases}$$

Exercise 3 (transformation and convergence in distribution). Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of RVRVs with $X_n \sim \mathrm{Unif}(0, 1/n)$ for each $n \in \mathbb{N}_{>0}$. Let $Y_n = nX_n$.

(iii) Determine the limiting cdf $F_Y(y) = \lim_{n\to\infty} F_{Y_n}(y)$.

$$F_{Y_n}(y) = \begin{cases} 0 & \text{if } y < 0 \\ y & \text{if } y \in [0,1] \\ 1 & \text{if } y > 1 \end{cases} \quad ; \quad F_Y(y) = \lim_{n\to\infty} F_{Y_n}(y) = F_{Y_n}(y)$$

Exercise 3 (transformation and convergence in distribution)**.** Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of RVRVs with $X_n \sim \mathrm{Unif}(0, 1/n)$ for each $n \in \mathbb{N}_{>0}$. Let $Y_n = nX_n$.

(iv) To what RVRV Y does Y_n converge in distribution? Specify its distribution.

$F_Y(y) = \lim_{n\to\infty} F_{Y_n}(y) = F_{Y_n}(y)$; $\quad Y_n \sim \mathrm{Unif}(0,1) \Rightarrow Y \sim \mathrm{Unif}(0,1)$

(iii) *converges in distribution* (or *converges weakly*) to X, denoted $X_n \xrightarrow{d} X$ or $X_n \Rightarrow X$, if

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

for every $x \in \mathbb{R}$ at which F_X is continuous. This means that the overall shape of the distribution of X_n approaches the shape of the distribution of X. Equivalently, $\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$ for all $t \in \mathbb{R}$.

Exercise 4 (mle for uniform distribution). Let $X_1, \dots, X_n$ be iid RVRVs from a uniform distribution $\mathrm{Unif}(0, \theta)$ with unknown parameter $\theta > 0$. The pdf of X_i is $p(x_i \mid \theta) = \frac{1}{\theta}\chi_{[0,\theta]}(x_i)$.

(i) Write down the likelihood function $L(\theta \mid x)$ for an observed sample $x = (x_1, \dots, x_n)$. Be careful with the indicator functions.

$X_1 \sim \mathrm{Unif}(0, \theta)$

$\rightarrow p_{X_i}(x) = \chi_{(0,\theta)}(x) \cdot \frac{1}{\theta}$

Definition 3.13 (likelihood function). Let $(X_1, \dots, X_n)$ be a random sample from a statistical model $(\Omega, \mathcal{A}, P_\Theta)$ where all $P_\theta \in P_\Theta$ have a pmf or density $p_\theta(\cdot) =: p(\cdot \mid \theta)$. Given observed data $x = (x_1, \dots, x_n)$, the *likelihood (function)* is the following function

$$L(\cdot \mid x) : \Theta \to \mathbb{R}_{\geq 0} : \theta \mapsto L(\theta \mid x) := p_\theta(x) = p(x \mid \theta) = p(x_1, \dots, x_n \mid \theta)\,.$$

If $X_1, \dots, X_n$ are iid, then $L(\theta \mid x) = \prod_{i=1}^n p(x_i \mid \theta)$. The *log-likelihood function* is $\ell(\theta \mid x) := \ln(L(\theta \mid x))$.

$x_{min} = \min\{x_1, \dots, x_n\}$

$x_{max} = \max\{x_1, \dots, x_n\}$

$$L(\theta \mid \vec{x}) = \prod_{i=1}^n p(x_i \mid \theta) = \prod_{i=1}^n \chi_{(0,\theta)}(x_i) \cdot \frac{1}{\theta} = \frac{1}{\theta^n} \cdot \underbrace{\prod_{i=1}^n \chi_{(0,\theta)}(x_i)}_{} = \frac{1}{\theta^n} \cdot \chi_{(x_{max}, \infty)}(\theta)$$

$$= \begin{cases} 1 & \text{if } (0 \leq x_i \leq \theta \mid \forall i \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} 1 & \text{if } 0 \leq x_{min} \leq x_{max} \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

$$= \chi_{(x_{max}, \infty)}(\theta)$$

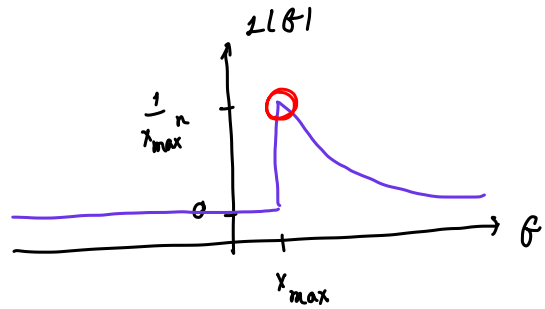
Exercise 4 (mle for uniform distribution). Let $X_1, \ldots, X_n$ be iid RVRVs from a uniform distribution $\mathrm{Unif}(0, \theta)$ with unknown parameter $\theta > 0$. The pdf of X_i is $p(x_i \mid \theta) = \frac{1}{\theta}\chi_{[0,\theta]}(x_i)$.

(ii) Find the maximum likelihood estimator (mle) $\hat{\theta}_{\mathrm{mle}}$ for θ.
[Hint: The likelihood function may not be differentiable with respect to θ, but the maximum value can be found by reasoning about monotonicity and where it is non-zero.]

Definition 3.27 (maximum likelihood estimation (mle)). The *maximum likelihood estimator (mle)* $\hat{\theta}_{\mathrm{mle}}$ is the value of $\theta \in \Theta$ that maximizes the likelihood function $L(\theta \mid x)$, i.e.,

$$\hat{\theta}_{\mathrm{mle}}(x) \in \arg\max_{\theta \in \Theta} L(\theta \mid x) = \arg\max_{\theta \in \Theta} \ell(\theta \mid x) .$$

Maximizing the log-likelihood is often mathematically simpler. If $\ell(\theta \mid x)$ is differentiable and the maximum is in the interior of Θ, $\hat{\theta}_{\mathrm{mle}}$ can often be found by solving the *likelihood equation* $\nabla_\theta \ell(\theta \mid x) = 0$ for θ.



$$L(\theta \mid \vec{x}) = \frac{1}{\theta^n} \cdot \chi_{(x_{max}, \infty)}(\theta)$$

$$\hat{\theta}_{mle}(x) = \arg\max_{\theta} \frac{1}{\theta^n} \cdot \chi_{(x_{max}, \infty)}(\theta)$$

$$\left(\frac{1}{\theta^n} \cdot \mathbb{1}\{\theta \geq x_{max}\} \right)$$

$$\hat{\theta}_{mle} = X_{max} = \max\{X_1, \ldots, X_n\}$$

Exercise 4 (mle for uniform distribution). Let $X_1, \ldots, X_n$ be iid RVRVs from a uniform distribution $\mathrm{Unif}(0, \theta)$ with unknown parameter $\theta > 0$. The pdf of X_i is $p(x_i \mid \theta) = \frac{1}{\theta}\chi_{[0,\theta]}(x_i)$.

(iii) Calculate the expected value $\mathbb{E}[\hat{\theta}_{\mathrm{mle}}]$. Is $\hat{\theta}_{\mathrm{mle}}$ an unbiased estimator for θ?
[Hint: First find the cdf of $\hat{\theta}_{\mathrm{mle}}$, then its pdf, and then compute the expectation.]

$\hat{\theta}_{mle} = X_{max}$

Definition 3.20 (properties of estimators). Let $\hat{\theta}$ be an estimator for the parameter θ. The quality of this estimator can be assessed according to various properties.

(i) **Bias:** The *bias* of an estimator $\hat{\theta}$ is $\mathrm{bias}[\hat{\theta}] = \mathbb{E}[\hat{\theta}] - \theta$. An estimator is *unbiased* if $\mathrm{bias}[\hat{\theta}] = 0$, i.e., $\mathbb{E}[\hat{\theta}] = \theta$ for all $\theta \in \Theta$. (What is the expectation over?)

(ii) **Consistency:** An estimator $\hat{\theta}_n$ is *consistent* if $\hat{\theta}_n \xrightarrow{P} \theta$. It is *strongly consistent* if $\hat{\theta}_n \xrightarrow{a.s.} \theta$.

(iii) **Variance:** The variance of an estimator is $\mathrm{var}[\hat{\theta}] = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$. For multi-variate θ, this turns into $\mathrm{cov}[\hat{\theta}, \hat{\theta}]$.

$\mathbb{E}[X_{max}]$

Exercise 4 (mle for uniform distribution). Let $X_1, \ldots, X_n$ be iid RVRVs from a uniform distribution $\mathrm{Unif}(0, \theta)$ with unknown parameter $\theta > 0$. The pdf of X_i is $p(x_i \mid \theta) = \frac{1}{\theta}\chi_{[0,\theta]}(x_i)$.

(iii) Calculate the expected value $\mathbb{E}[\hat{\theta}_{\mathrm{mle}}]$. Is $\hat{\theta}_{\mathrm{mle}}$ an unbiased estimator for θ?
[Hint: First find the cdf of $\hat{\theta}_{\mathrm{mle}}$, then its pdf, and then compute the expectation.]

$$\hat{\theta}_{mle} = X_{max}$$

$$F_{X_{max}}(t) = P(X_{max} \le t) = P(X_1 \le t, X_2 \le t, \ldots, X_n \le t) = \prod_{j=1}^{n} P(X_j \le t)$$

$$= \prod_{j=1}^{n} \begin{cases} 0 & \text{if } t < 0 \\ t/\theta & \text{if } t \in [0, \theta] \\ 1 & \text{if } t \geq \theta \end{cases}$$

» For $t \in (0, \theta)$; $F_{X_{max}}(t) = (t/\theta)^n$

$$p_{X_{max}}(t) = \frac{d}{dt} F_{X_{max}}(t) = \begin{cases} n \cdot \dfrac{t^{n-1}}{\theta^n} & \text{if } 0 \le t \le \theta \\ 0 & \text{otherwise} \end{cases}$$

Exercise 4 (mle for uniform distribution). Let $X_1, \ldots, X_n$ be iid RVRVs from a uniform distribution $\mathrm{Unif}(0, \theta)$ with unknown parameter $\theta > 0$. The pdf of X_i is $p(x_i \mid \theta) = \frac{1}{\theta}\chi_{[0,\theta]}(x_i)$.

(iii) Calculate the expected value $\mathbb{E}[\hat{\theta}_{\mathrm{mle}}]$. Is $\hat{\theta}_{\mathrm{mle}}$ an unbiased estimator for θ?
[Hint: First find the cdf of $\hat{\theta}_{\mathrm{mle}}$, then its pdf, and then compute the expectation.]

(Def 3.20)

$$p_{X_{max}}(t) = \frac{d}{dt} F_{X_{max}}(t) = \begin{cases} n \cdot \frac{t^{n-1}}{\theta^n} & \text{if } 0 \leq t \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[X_{max}] = \int_0^\theta t \cdot \frac{n \cdot t^{n-1}}{\theta^n} \, dt = \cdots = \underbrace{\frac{n}{n+1}}_{\substack{\to 1 \\ \text{as} \\ n \to \infty}} \cdot \theta \neq \theta \text{ exactly} \Rightarrow \text{biased}$$

$\Rightarrow$ asymptotically unbiased

Definition 2.46 (characteristic function (cf)). Let X be an RVRV. The *characteristic function (cf)* of X, denoted $\phi_X(t)$ or $\varphi_X(t)$, is defined as

$$\phi_X(t) = \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX) + i\sin(tX)]\,,$$

where $i = \sqrt{-1}$ is the imaginary unit.

(i) If X is discrete with pmf $p_X(x)$, $\phi_X(t) = \sum_x e^{itx} p_X(x)$.

(ii) If X is continuous with pdf $p_X(x)$, $\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} p_X(x)\,dx$.

Remark 2.47 (properties of cfs). Let X, Y be RVRVs.

(i) $\phi_X(0) = 1$ and $|\phi_X(t)| \leq 1$.

(ii) $\phi_{aX+b}(t) = e^{itb}\phi_X(at)$.

(iii) We can compute *moments* from the derivatives of $\phi(t)$ via $\mathbb{E}[X^n] = i^{-n}\phi_X^{(n)}(0)$ (if the moments exist).

(iv) If $X \perp\!\!\!\perp Y$, then $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.

(v) **Uniqueness:** The cf uniquely determines the distribution. There are explicit inversion formulas to recover the cdf or pdf from the cf (e.g., Lévy's inversion theorem).

(iv) *Exponential distribution:* The special case $\mathrm{Exp}(\lambda) = \Gamma(1, \lambda)$ with $\lambda > 0$ on $\Omega = \mathbb{R}_{>0}$ is characterized via the pdf

$$p_{\mathrm{Exp}(\lambda)} : \mathbb{R} \to \mathbb{R}_{\geq 0},\ x \mapsto \chi_\Omega(x)\, \lambda e^{-\lambda x}\,.$$

It models the time until the first event in a Poisson process of rate λ, for example the time until the next bus arrives or the length of a phone call in a call center.

Theorem 2.40 (Chebyshev's inequality). *Let X be an RVRV with finite mean $\mathbb{E}[X]$ and finite variance $\mathrm{var}[X] > 0$. Then for any $\epsilon > 0$,*

$$P(|X - \mathbb{E}[X]| \geq \epsilon) \leq \frac{\mathrm{var}[X]}{\epsilon^2}\,.$$

Definition 2.56 (modes of convergence). Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of RVRV and let X be another RVRV (which can also be a constant), all defined on the same probability space $(\Omega, \mathcal{A}, P)$. The sequence $(X_n)_{n\in\mathbb{N}}$

(i) *converges in probability* to X, denoted $X_n \xrightarrow{P} X$, if for every $\epsilon > 0$,

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

Intuitively, this means that for large n, it is very unlikely that X_n is far from X.

(ii) *converges almost surely* to X, denoted $X_n \xrightarrow{a.s.} X$ or $X_n \to X$ a.s., if

$$P\left(\left\{\omega \in \Omega \,\middle|\, \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1\,.$$

Intuitively, this means that for almost all outcomes ω, the sequence of numbers $X_n(\omega)$ converges to $X(\omega)$.

(iii) *converges in distribution* (or *converges weakly*) to X, denoted $X_n \xrightarrow{d} X$ or $X_n \Rightarrow X$, if

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

for every $x \in \mathbb{R}$ at which F_X is continuous. This means that the overall shape of the distribution of X_n approaches the shape of the distribution of X. Equivalently, $\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$ for all $t \in \mathbb{R}$.

Remark 2.57 (relations between modes of convergence). The different modes of convergence are related via

$$X_n \xrightarrow{a.s.} X \quad \Longrightarrow \quad X_n \xrightarrow{P} X \quad \Longrightarrow \quad X_n \xrightarrow{d} X\,.$$

In words,

$$\text{almost sure conv.} \Longrightarrow \text{conv. in probability} \Longrightarrow \text{conv. in distribution}\,.$$

Theorem 2.64 (central limit theorem (clt)). *Let $X_1, X_2, \ldots$ be a sequence of iid RVRVs with finite mean $\mathbb{E}[X_i] = \mu$ and finite, non-zero variance $\operatorname{var}[X_i] = \sigma^2 \in \mathbb{R}_{>0}$. Let $S_n := \sum_{i=1}^n X_i$ be the sum, and $\bar{X}_n = S_n/n$ be the sample mean. Consider the* standardized sum *(or* standardized mean*):*

$$Z_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\operatorname{var}[S_n]}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} .$$

Then

$$Z_n \xrightarrow{d} \mathcal{N}(0,1) .$$

The quantity $\sigma/\sqrt{n}$ that dictates the "asymptotic standard deviation" of the normal distribution as n goes to infinity is called standard error*.*

Definition 3.13 (likelihood function). Let $(X_1, \ldots, X_n)$ be a random sample from a statistical model $(\Omega, \mathcal{A}, P_\Theta)$ where all $P_\theta \in P_\Theta$ have a pmf or density $p_\theta(\cdot) =: p(\cdot \mid \theta)$. Given observed data $x = (x_1, \ldots, x_n)$, the *likelihood (function)* is the following function

$$L(\cdot \mid x) : \Theta \to \mathbb{R}_{\geq 0} : \theta \mapsto L(\theta \mid x) := p_\theta(x) = p(x \mid \theta) = p(x_1, \ldots, x_n \mid \theta) .$$

If $X_1, \ldots, X_n$ are iid, then $L(\theta \mid x) = \prod_{i=1}^n p(x_i \mid \theta)$. The *log-likelihood function* is $\ell(\theta \mid x) := \ln(L(\theta \mid x))$.

Definition 3.20 (properties of estimators). Let $\hat{\theta}$ be an estimator for the parameter θ. The quality of this estimator can be assessed according to various properties.

(i) **Bias:** The *bias* of an estimator $\hat{\theta}$ is $\text{bias}[\hat{\theta}] = \mathbb{E}[\hat{\theta}] - \theta$. An estimator is *unbiased* if $\text{bias}[\hat{\theta}] = 0$, i.e., $\mathbb{E}[\hat{\theta}] = \theta$ for all $\theta \in \Theta$. (What is the expectation over?)

(ii) **Consistency:** An estimator $\hat{\theta}_n$ is *consistent* if $\hat{\theta}_n \xrightarrow{P} \theta$. It is *strongly consistent* if $\hat{\theta}_n \xrightarrow{a.s.} \theta$.

(iii) **Variance:** The variance of an estimator is $\text{var}[\hat{\theta}] = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$. For multi-variate θ, this turns into $\text{cov}[\hat{\theta}, \hat{\theta}]$.

(v) **Asymptotic normality:** An estimator $\hat{\theta}_n$ is *asymptotically normal* if its distribution, when properly centered and scaled, converges to a normal distribution. Formally, if there exists a quantity $V(\theta)$ such that

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, V(\theta)) \,.$$

The term $V(\theta)$ is the *asymptotic variance*. This property is fundamental for constructing approximate confidence intervals and hypothesis tests for large sample sizes, as it implies $\hat{\theta}_n$ approximately follows $\mathcal{N}(\theta, V(\theta)/n)$.

(vi) **Efficiency:** Efficiency compares the variance of unbiased estimators. An estimator is a *minimum variance unbiased estimator (mvue)* if it achieves the lowest possible variance among all unbiased estimators.

Example 3.21 (properties of sample mean). Let $X_1, \ldots, X_n$ be iid RVRVs with mean μ and finite variance $\sigma^2 > 0$. More precisely, assume $X_1, \ldots, X_n \sim P_\theta$ iid, where $\theta = (\mu, \sigma^2, \ldots) \in \Theta$ is the true parameter in our statistical model and μ is the mean of P_θ and σ^2 is the variance of P_θ. Note that neither P_θ (nor any other distribution in our model) need to be Gaussian. There could be more parameters in θ describing other properties of the distribution P_θ as well.

Consider the estimator $\hat{\mu} = \bar{X}_n = \frac{1}{n} \sum X_i$ for the parameter of interest $\gamma(\theta) = \mu$.

(i) The sample mean $\bar{X}_n$ is an unbiased estimator of μ, $\mathbb{E}[\bar{X}_n] = \mu$.

(ii) By the wlln and slln, $\bar{X}_n \xrightarrow{P} \mu$ and $\bar{X}_n \xrightarrow{a.s.} \mu$. So $\bar{X}_n$ is strongly consistent.

(iii) The variance decreases in n as $\mathrm{var}[\bar{X}_n] = \sigma^2/n$. Note that this is also what we would expect from the clt.

(iv) Since $\hat{\mu}$ is unbiased, $\mathrm{mse}[\bar{X}_n] = \mathrm{var}[\bar{X}_n] = \sigma^2/n$.

(v) By the clt, the sample mean $\bar{X}_n$ is an asymptotically normal estimator with asymptotic variance $V(\theta) = \sigma^2$.

(vi) The sample mean $\bar{X}_n$ is efficient for Gaussian models, but need not be efficient for other distributions.

Definition 3.27 (maximum likelihood estimation (mle)). The *maximum likelihood estimator (mle)* $\hat{\theta}_{\text{mle}}$ is the value of $\theta \in \Theta$ that maximizes the likelihood function $L(\theta \mid x)$, i.e.,

$$\hat{\theta}_{\text{mle}}(x) \in \arg\max_{\theta \in \Theta} L(\theta \mid x) = \arg\max_{\theta \in \Theta} \ell(\theta \mid x) \,.$$

Maximizing the log-likelihood is often mathematically simpler. If $\ell(\theta \mid x)$ is differentiable and the maximum is in the interior of Θ, $\hat{\theta}_{\text{mle}}$ can often be found by solving the *likelihood equation* $\nabla_\theta \ell(\theta \mid x) = 0$ for θ.