

Math 4261: Mathematical Statistics I

Lecture 10: Review 1

Fall 2026

Outline

Probability tools

Best and best linear estimates

Order statistics and quantile estimation

Method of moments

Outline

Probability tools

Best and best linear estimates

Order statistics and quantile estimation

Method of moments

Standard Gaussian

Let $Z \sim \mathcal{N}(0, 1)$. Its **CDF** is often denoted

$$\Phi(x) = \mathbb{P}[Z \leq x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt.$$

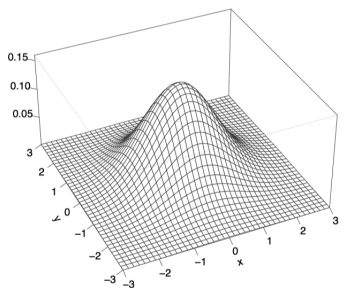
This has **no closed form!** Precise numerical approximations are available.

Symmetry about zero For $x > 0$,

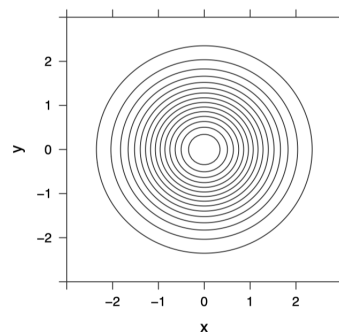
- ▶ $\mathbb{P}[Z \leq -x] = \Phi(-x) = \mathbb{P}[Z \geq x]$
- ▶ $\mathbb{P}[|Z| \leq x] = 2\Phi(x) - 1$

Standard multivariate Gaussian

$$\mathbf{Z} = (Z_1, \dots, Z_p)^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p \times p}), \quad \text{i.e. } Z_1, \dots, Z_p \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1).$$



PDF surface



Circular contours

- ▶ PDF $f(z_1, \dots, z_p) = \frac{1}{(2\pi)^{p/2}} \exp \left\{ -\frac{1}{2} \|\mathbf{z}\|^2 \right\}$
- ▶ The distribution is **spherically symmetric**: rotations and reflections do not change it: $Q\mathbf{Z} \stackrel{d}{=} \mathbf{Z}$ for any orthonormal Q .
- ▶ $\mathbf{a}^\top \mathbf{Z} \sim \mathcal{N}(0, \|\mathbf{a}\|^2)$. If $\|\mathbf{a}\| = 1$ then $\mathbf{a}^\top \mathbf{Z} \sim \mathcal{N}(0, 1)$.

Moment generating functions

The **moment generating function** is $M_X(t) = \mathbb{E}[e^{tX}]$. If it is finite on an interval around 0, it determines the distribution.

For independent $X_1, \dots, X_n$, and constants a, b ,

$$M_{X_1 + \dots + X_n}(t) = \prod_{i=1}^n M_{X_i}(t), \quad M_{aX+b}(t) = e^{bt} M_X(at).$$

Examples

$$X \sim \text{Bern}(p) : \quad M_X(t) = 1 - p + pe^t,$$

$$X \sim \text{Bin}(n, p) : \quad M_X(t) = (1 - p + pe^t)^n,$$

$$Z \sim \mathcal{N}(0, 1) : \quad M_Z(t) = e^{t^2/2},$$

$$Y \sim \chi_n^2 : \quad M_Y(t) = (1 - 2t)^{-n/2}, \quad t < 1/2.$$

Applications of MGF

The MGF *generates* all the moments of X :

Theorem

If $M_X(t) < \infty$ in an interval around $t = 0$, then

$$\mathbb{E}[X^k] = \left. \frac{d^k}{dt^k} M_X(t) \right|_{t=0} = M_X^{(k)}(0).$$

The MGF of a random variable **determines its distribution**:

Theorem

Let X and Y be two r.v.'s such that $M_X(t) = M_Y(t) < \infty$ in an interval around $t = 0$. Then they have the same distribution.

Examples:

- ▶ If X is **Bernoulli** distributed, then $\mathbb{E}[X] = p$ determines the distribution.
- ▶ If X is **Gaussian** distributed, then $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$ determine the distribution (variance is a function of the first two moments).

Deviation inequalities/tail bounds

Markov's inequality

Let X be a **non-negative** random variable and $a > 0$. Then

$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

Chebyshev's inequality

Let X have finite variance and $d > 0$. Then

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq d] \leq \frac{\text{Var}[X]}{d^2}.$$

Gaussian tail bound

Let $Z \sim \mathcal{N}(0, 1)$ and $t \geq 0$. Then

$$\mathbb{P}[|Z| \geq t] \leq 2e^{-t^2/2}.$$

Modes of convergence

Convergence in probability

$$T_n \xrightarrow{p} T \iff \mathbb{P}[|T_n - T| > \epsilon] \longrightarrow 0 \quad \text{for every } \epsilon > 0.$$

Convergence in distribution

$$T_n \xrightarrow{d} T \iff F_{T_n}(x) \longrightarrow F_T(x)$$

at every x where F_T is continuous.

- ▶ $T_n \xrightarrow{p} T$ implies $T_n \xrightarrow{d} T$. Converse false in general.
- ▶ For a **constant** c , $T_n \xrightarrow{d} c$ is equivalent to $T_n \xrightarrow{p} c$.

Law of large numbers

Let $X_1, \dots, X_n$ be i.i.d., with mean μ and variance $\sigma^2 < \infty$.

Law of large numbers (LLN)

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu.$$

Statistical implication: The sample mean is a **consistent estimator** of the population mean (under broad conditions).

Central limit theorem

Let $X_1, \dots, X_n$ be i.i.d., with mean μ and variance $0 < \sigma^2 < \infty$.

Central limit theorem (CLT)

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Equivalently,

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

For large n , $\bar{X}_n$ is approximately $\mathcal{N}(\mu, \sigma^2/n)$ distributed.

The LLN says the sampling error vanishes. The CLT describes its approximately Gaussian shape and its $1/\sqrt{n}$ scale.

Continuous mapping theorem

Often we might know the convergence of some statistic T_n , but we want to understand the convergence of $g(T_n)$ for some function g . A useful tool is

Continuous mapping theorem

Let $g(x)$ be a continuous function of x . As $n \rightarrow \infty$,

(a) If $T_n \xrightarrow{d} Z$, then $g(T_n) \xrightarrow{d} g(Z)$.

(b) If $T_n \xrightarrow{p} Z$, then $g(T_n) \xrightarrow{p} g(Z)$.

Example: consistency of the empirical variance

Assuming additionally $\mathbb{E}[X_1^4] < \infty$, we showed

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 \xrightarrow{p} \sigma^2.$$

Slutsky's lemma

Slutsky's lemma

Suppose $T_n \xrightarrow{d} T$ and $A_n \xrightarrow{d} a$, where a is a constant. Then

$$T_n + A_n \xrightarrow{d} T + a,$$

$$A_n T_n \xrightarrow{d} aT,$$

$$\frac{T_n}{A_n} \xrightarrow{d} \frac{T}{a} \quad \text{if } a \neq 0.$$

- ▶ Independence is not required between T_n and A_n
- ▶ Convergence in distribution is **not** generally closed under sum or product.

Conditional independence

Consider random variables X , Y , and Z

Independence of X and Y

Knowing Y **does not change** the distribution of X .

Conditional independence given Z

Once Z is known, learning Y does not change the distribution of X (and vice versa).

For each fixed value of Z , the r.v.'s X and Y are independent under the conditional distribution.

Neither property implies the other.

- ▶ Variables can be independent given Z but dependent overall.
- ▶ Independent variables can become dependent after conditioning on Z .

Example: Review 1, Q1

We have two coins: a fair coin and a coin whose probability of heads is $3/4$. **One coin** is selected at random, with each coin equally likely, and the selected coin is tossed repeatedly. Let $\Theta \in \{1/2, 3/4\}$ be the probability of heads for the selected coin, and let X_i be the indicator that the i th toss is heads. Conditional on Θ , the tosses are independent and each X_i has the Bernoulli(Θ) distribution.

(a) Find the probabilities $\mathbb{P}[X_i = 1]$ and $\mathbb{P}[X_1 = 1, X_2 = 1]$. Are X_1 and X_2 independent?

$$\begin{aligned} \mathbb{P}[X_i = 1] &\stackrel{\text{Law total prob}}{=} \underbrace{\mathbb{P}[X_i = 1 \mid \Theta = \frac{1}{2}]}_{\frac{1}{2}} \underbrace{\mathbb{P}[\Theta = \frac{1}{2}]}_{\frac{1}{2}} \\ &\quad + \underbrace{\mathbb{P}[X_i = 1 \mid \Theta = \frac{3}{4}]}_{\frac{3}{4}} \underbrace{\mathbb{P}[\Theta = \frac{3}{4}]}_{\frac{1}{2}} = \frac{5}{8} \quad \text{holds } \forall i \end{aligned}$$

$$\begin{aligned} \mathbb{P}[X_1 = 1, X_2 = 1] &\stackrel{\text{Law total prob}}{=} \underbrace{\mathbb{P}[X_1 = 1, X_2 = 1 \mid \Theta = \frac{1}{2}]}_{\substack{\text{Cond independence} \\ (\frac{1}{2})^2}} \frac{1}{2} + \underbrace{\mathbb{P}[X_1 = 1, X_2 = 1 \mid \Theta = \frac{3}{4}]}_{(\frac{3}{4})^2} \frac{1}{2} \\ &= \frac{13}{32} \end{aligned}$$

Example: Review 1, Q1

We have two coins: a fair coin and a coin whose probability of heads is $3/4$. **One coin** is selected at random, with each coin equally likely, and the selected coin is tossed repeatedly. Let $\Theta \in \{1/2, 3/4\}$ be the probability of heads for the selected coin, and let X_i be the indicator that the i th toss is heads. Conditional on Θ , the tosses are independent and each X_i has the $\text{Bernoulli}(\Theta)$ distribution.

(a) Find the probabilities $\mathbb{P}[X_i = 1]$ and $\mathbb{P}[X_1 = 1, X_2 = 1]$. Are X_1 and X_2 independent?

$$\mathbb{P}[X_i = 1] = \frac{5}{8} \quad \begin{array}{l} \text{holds} \\ \forall i \end{array}, \quad \mathbb{P}[X_1 = 1, X_2 = 1] = \frac{13}{32}$$

$$\frac{13}{32} = \mathbb{P}[X_1 = 1, X_2 = 1] \neq \mathbb{P}[X_1 = 1] \mathbb{P}[X_2 = 1] = \left(\frac{5}{8}\right)^2 = \frac{25}{64}$$

$$\begin{array}{r} \uparrow \\ 26 \\ \hline 64 \end{array}$$

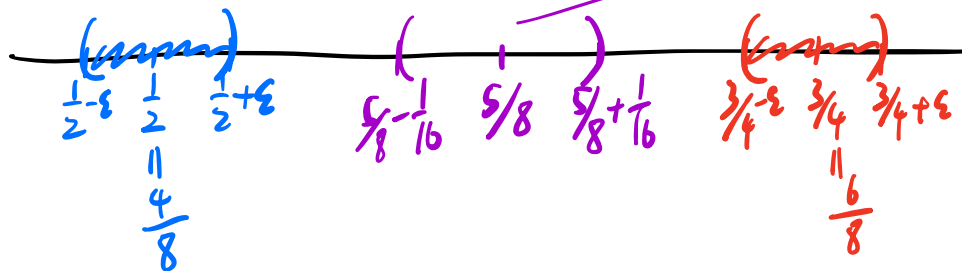
Example: Review 1, Q1

Let $\Theta \in \{1/2, 3/4\}$: probability of heads for the selected coin. Let X_i be the indicator that the i th toss is heads and $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

(b) Apply LLN conditional on the selected coin. What does $\bar{X}_n$ approach for each possible choice of coin? Does $\bar{X}_n$ converge in probability to $5/8$?

$$\bar{X}_n \xrightarrow{P} \begin{cases} \frac{1}{2}, & \text{Cond. on } \Theta = \frac{1}{2} \\ \frac{3}{4}, & \text{Cond. on } \Theta = \frac{3}{4} \end{cases}$$

$$\text{i.e. } P\left[\underbrace{|\bar{X}_n - \frac{1}{2}| > \varepsilon}_{\bar{X}_n \notin (\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon)} \mid \Theta = \frac{1}{2}\right] \rightarrow 0 \quad \forall \varepsilon$$



$\bar{X}_n$ lies here
with vanishing
probability
so $\bar{X}_n \not\rightarrow \frac{5}{8}$

Example: Review 1, Q1

(c) Now change the experiment: **before every toss**, independently select one of the two coins with equal probabilities and toss it once. What does the observed proportion of heads converge to in probability under this procedure?

$$\text{Now } X_i \stackrel{\text{iid}}{\sim} \text{Ber}\left(\frac{5}{8}\right)$$

By LLN

$$\overline{X}_n \xrightarrow{P} \frac{5}{8}$$

Outline

Probability tools

Best and best linear estimates

Order statistics and quantile estimation

Method of moments

Conditional expectation

The **conditional expectation** of Y given $X = x$ is the **number**

$$\mathbb{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x) \, dy = g(x)$$

(replace the integral by a sum and the conditional PDF by a PMF if discrete). The conditional expectation of Y given X is the **random variable**

$$\mathbb{E}[Y \mid X] = g(X).$$

Application: Law of total expectation

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$$

Example: Conditional expectation

(Revisit Review 1 Q1): one coin has probability of heads $1/2$, the other $3/4$. Select one coin at random and toss it five times, independently given the selected coin. Let Θ be its probability of heads and Y the **number of heads**.

(i) Given $\Theta = 1/2$, we are tossing the **fair coin** five times.

$$\mathbb{E}[Y \mid \Theta = 1/2] = 1 \cdot p = 5 \cdot \frac{1}{2}$$

(ii) Given $\Theta = 3/4$, we are tossing the **biased coin** five times.

$$\mathbb{E}[Y \mid \Theta = 3/4] = 5 \cdot \frac{3}{4}.$$

Each answer is a **number**. The **random variable** $\mathbb{E}[Y \mid \Theta]$ takes the answer corresponding to the selected coin.

Conditional variance

The **conditional variance** of Y given X is the **random variable**

$$\text{Var}[Y \mid X] = \mathbb{E}[Y^2 \mid X] - (\mathbb{E}[Y \mid X])^2$$

Application: Eve's law (E-V-V-E)

$$\text{Var}[Y] = \mathbb{E}[\text{Var}[Y \mid X]] + \text{Var}[\mathbb{E}[Y \mid X]]$$

Best estimates

Let X be observed data and Y the quantity of interest to predict. We **choose** an estimator $\delta(X)$ to minimize

$$\text{MSE}(\delta) = \mathbb{E}[(Y - \delta(X))^2].$$

The **best estimator** is

$$\delta_{\text{opt}}(X) = \mathbb{E}[Y \mid X].$$

Its MSE is $\mathbb{E}[\text{Var}[Y \mid X]]$.

If we restrict to **constant estimators**, the best estimator is the mean:
 $c_{\text{opt}} = \mathbb{E}[Y]$.

The **best linear estimator** is

$$\delta_{\text{lin}}(X) = \mathbb{E}[Y] + \frac{\text{Cov}[X, Y]}{\text{Var}[X]}(X - \mathbb{E}[X]).$$

Motivation

The best estimator for predicting Y given X (in the MSE sense) is the conditional expectation $\mathbb{E}[Y \mid X]$.

A **linear estimator** uses

$$aX + b, \quad \text{for some } a, b \in \mathbb{R}$$

to predict Y .

Two reasons why linear is desirable:

- ▶ Linear rule is simple and interpretable
- ▶ Computing best rule $\mathbb{E}[Y \mid X]$ usually requires the conditional or joint law, which might not be available.

It turns out for best linear estimate, we only need **mean, variance, and covariance!**

Best linear estimator

For $\text{Var}[X] > 0$, the best linear estimator $\delta(X) = aX + b$ has

$$\text{Best coefficients} = \begin{cases} a = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} \\ b = \mathbb{E}[Y] - a\mathbb{E}[X] \end{cases}$$

Thus,

$$\text{Best linear estimator} = aX + b = \boxed{\mathbb{E}[Y] + \frac{\text{Cov}[X, Y]}{\text{Var}[X]}(X - \mathbb{E}[X])}$$

This achieves the minimum MSE among all linear estimators:

$$(1 - \rho^2) \text{Var}[Y] = \text{Var}[Y] - \frac{\text{Cov}[X, Y]^2}{\text{Var}[X]}$$

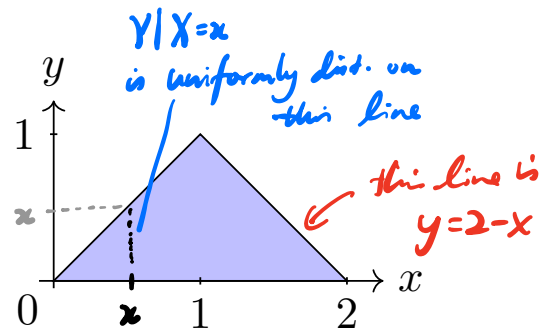
Example: Review 1, Q3

Let (X, Y) be uniformly distributed over the triangle with vertices $(0, 0)$, $(2, 0)$, and $(1, 1)$. We want to estimate Y from X under squared-error loss.

Write $h(x) = \min\{x, 2 - x\}$ for $0 < x < 2$. You may use

$$Y \mid X = x \sim \text{Unif}(0, h(x)),$$

$$\mathbb{E}[h(X)^2] = \frac{1}{2}.$$



(i) Find the **best estimate** of Y as a function of X , and its MSE.

$$\mathbb{E}[Y \mid X = x] = \frac{h(x)}{2} \text{ (midpoint) } \text{ i.e. } \mathbb{E}[Y \mid X] = \frac{h(X)}{2}$$

$$\text{MSE} = \mathbb{E}[(Y - \mathbb{E}[Y \mid X])^2] \stackrel{\text{Law of total expec.}}{=} \mathbb{E}[\underbrace{\mathbb{E}[(Y - \mathbb{E}[Y \mid X])^2 \mid X]}_{\text{Var}[Y \mid X]}]$$

$$\begin{aligned} \text{Var}[\text{Unif}(a,b)] &= \frac{(b-a)^2}{12} \\ &= \mathbb{E}\left[\frac{h(X)^2}{12}\right] = \frac{1}{24} \end{aligned}$$

Example: Review 1, Q3

We may compute:

$$\mathbb{E}[X] = 1, \quad \text{Var}[X] = \frac{1}{6}, \quad \mathbb{E}[Y] = \frac{1}{3}, \quad \text{Var}[Y] = \frac{1}{18},$$

$$\text{Cov}[X, Y] = 0.$$

(ii) Find the best **linear** estimate $aX + b$, and its MSE.

$$\mathcal{J}_{\text{lin}}(Y) = \frac{1}{3}$$

*(X has no linear relationship,
hence no linear predictive
power for Y)*

$$\text{MSE} = \mathbb{E}\left[\left(Y - \frac{1}{3}\right)^2\right] = \text{Var } Y = \frac{1}{18}$$

Example: Review 1, Q3

(iii) Are X and Y independent? Explain. Why can X help predict Y even though $\text{Cov}[X, Y] = 0$?

Dependent b/c cond. dist. of $Y|X=x$ depends on x

$$\text{Notice } \text{MSE}(\hat{Y}_{\text{best}}) = \frac{1}{24} < \text{MSE}(\hat{Y}_{\text{lin}}) = \frac{1}{18}.$$

So X can indeed be used to help predict Y .

Outline

Probability tools

Best and best linear estimates

Order statistics and quantile estimation

Method of moments

Order statistics

Let $X_1, \dots, X_n$ be random variables. Their **order statistics** are

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}.$$

Suppose X_i 's i.i.d. with CDF F . To find the distribution of $X_{(k)}$, consider

► $N_x = \sum_{i=1}^n \mathbf{1}\{X_i \leq x\} \sim \text{Bin}(n, F(x)).$

► $\{X_{(k)} \leq x\} = \{N_x \geq k\}$

Distribution of k -th order statistic

$$\mathbb{P}[X_{(k)} \leq x] = \sum_{j=k}^n \binom{n}{j} F(x)^j [1 - F(x)]^{n-j}.$$

Estimating quantiles

Aside from understanding the population mean and variance, we may want to learn a **population quantile**, such as the median or the 90th percentile.

Idea: estimate quantile using the corresponding **ranked observation** in the sample. I.e. order statistics.

Example: $\text{Uniform}(0, \theta)$

For $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Unif}(0, \theta)$ with $\theta > 0$, we estimated $q_1 = \theta$ by $X_{(n)}$. The sampling distribution is

$$\mathbb{P}[X_{(n)} \leq x] = \left(\frac{x}{\theta}\right)^n, \quad 0 \leq x \leq \theta.$$

This allows us to show that $X_{(n)}$ is consistent for θ .

Estimating quantiles

More generally, to estimate the population quantile for $0 < \alpha < 1$,¹

$$q_\alpha = F^{-1}(\alpha) = \inf\{x : F(x) \geq \alpha\},$$

we use

Plug-in quantile estimator

$$\hat{q}_{n,\alpha} = \inf\{x : \widehat{F}_n(x) \geq \alpha\} = X_{(\lceil n\alpha \rceil)}.$$

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

► Scanning from the left, $\hat{q}_{n,\alpha}$ is the first place $\widehat{F}_n$ reaches α .

E.g. For 1, 2, 2, 4, 7, 9, our sample median is $\hat{q}_{6,0.5} = X_{(3)} = 2$ (the lower median for even n).

Today: Is this estimator consistent? What can we say about its sampling distribution?

¹The $\alpha = 0, 1$ cases may require separate treatment.

Consistency of sample quantiles

Fix $0 < \alpha < 1$ and let $q_\alpha = F^{-1}(\alpha)$.

Quantile consistency

Suppose $X_1, \dots, X_n$ are i.i.d. and, for every $\epsilon > 0$,

$$F(q_\alpha - \epsilon) < \alpha < F(q_\alpha + \epsilon).$$

Then the empirical-inverse sample quantile satisfies

$$\hat{q}_{n,\alpha} \xrightarrow{p} q_\alpha.$$

- ▶ The strict inequalities let the LLN keep the empirical CDF below α on the left and above α on the right, **trapping its first crossing** between the two thresholds.
- ▶ A continuous, positive density near q_α is one sufficient condition.

A CLT for sample quantiles

Fix $0 < \alpha < 1$. Suppose F has a continuous density f near q_α , and that $f(q_\alpha) > 0$.

Asymptotic normality of a sample quantile

For an i.i.d. sample,

$$\sqrt{n}(\hat{q}_{n,\alpha} - q_\alpha) \xrightarrow{d} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f(q_\alpha)^2}\right).$$

Thus a large-sample approximation is

$$\hat{q}_{n,\alpha} \approx \mathcal{N}\left(q_\alpha, \frac{\alpha(1-\alpha)}{n f(q_\alpha)^2}\right).$$

Example: Exponential quantiles

Let $X_1, \dots, X_n$ be i.i.d. $\text{Expo}(1)$. Suppose we have the following ten observations, arranged in increasing order:

0.1, 0.2, 0.3, 0.5, 0.6, 0.9, 1.1, 1.4, 1.8, 2.5.

(i) Using our sample quantile convention, find the sample median $\hat{q}_{10,0.5}$ and the sample 75th percentile $\hat{q}_{10,0.75}$.

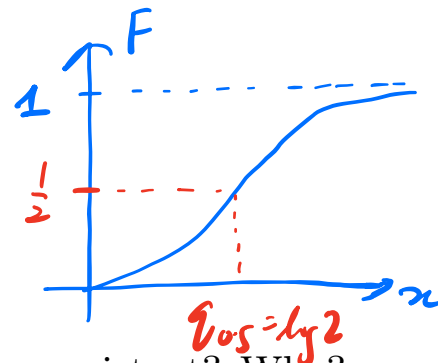
$$\hat{q}_{10,0.5} = X_{\lceil 10 \cdot \frac{1}{2} \rceil} = X_5 = 0.6$$

$$\hat{q}_{10,0.75} = X_{\lceil 10 \cdot \frac{3}{4} \rceil} = X_{\lceil 7.5 \rceil} = X_8 = 1.4$$

Example: Exponential quantiles

For $x \geq 0$, the CDF and density is

$$F(x) = 1 - e^{-x}, \quad f(x) = e^{-x}.$$



(ii) Find the population median $q_{0.5}$. Is $\hat{q}_{n,0.5} = X_{(\lceil n/2 \rceil)}$ consistent? Why?

$$\frac{1}{2} = P[X < M] = 1 - e^{-M}$$

$$\text{Solve: } e^{-M} = \frac{1}{2} \Leftrightarrow M = \ln 2$$

Yes, $\hat{q}_{n,0.5}$ is consistent - Because for any $0 < \alpha < 1$,

$$\text{for any } \varepsilon > 0, \quad F(q_\alpha - \varepsilon) < F(q_\alpha) < F(q_\alpha + \varepsilon)$$

(density is positive and continuous at such q_α)

hence theorem applies.

Example: Exponential quantiles

For $x \geq 0$, the CDF and density is

$$F(x) = 1 - e^{-x}, \quad f(x) = e^{-x}.$$

(iii) State the asymptotic distribution of $\sqrt{n}(\hat{q}_{n,0.5} - q_{0.5})$, giving the limiting variance explicitly.

$$f(\log 2) = \frac{1}{2}$$

$$\sqrt{n}(\hat{q}_{n,0.5} - q_{0.5}) \xrightarrow{d} \mathcal{N}\left(0, \frac{\frac{1}{2}(1-\frac{1}{2})}{f(q_{0.5})^2}\right) = \mathcal{N}(0, 1)$$

$$\text{i.e. } \hat{q}_{n,0.5} \sim \mathcal{N}(\log 2, \frac{1}{n}) \text{ approximately for large } n.$$

Outline

Probability tools

Best and best linear estimates

Order statistics and quantile estimation

Method of moments

Parameter estimation and bias

Model: $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} f(x \mid \theta)$, where $\theta \in \Omega \subseteq \mathbb{R}^k$ is unknown and Ω is the **parameter space**.

An **estimator** $\hat{\theta} = T(X_1, \dots, X_n)$ is a function of the data, so it is random. For a scalar parameter, its bias is

$$\text{Bias}_{\theta}(\hat{\theta}) = \mathbb{E}_{\theta}[\hat{\theta}] - \theta.$$

It is **unbiased** if $\mathbb{E}_{\theta}[\hat{\theta}] = \theta$ for **every** $\theta \in \Omega$.

The **variance** of $\hat{\theta}$

$$\text{Var}_{\theta}[\hat{\theta}] = \text{Var}_{\theta}[\hat{\theta}(X_1, \dots, X_n)].$$

The **standard error** of $\hat{\theta}$ is the standard deviation $\sqrt{\text{Var}_{\theta}[\hat{\theta}]}$.

Bias-Variance decomposition

The **mean-square error (MSE)** of an estimator $\hat{\theta}$ is

$$\mathbb{E}_{\theta}[(\hat{\theta} - \theta)^2]$$

Bias-Variance decomposition

$$\begin{aligned}\mathbf{MSE} &= \text{bias}^2 + \text{variance} \\ &= (\mathbb{E}_{\theta}[\hat{\theta}] - \theta)^2 + \mathbb{E}_{\theta}[(\hat{\theta} - \mathbb{E}_{\theta}[\hat{\theta}])^2]\end{aligned}$$

Bias, variance, and MSE generally depend on the true parameter value θ . The same estimator can be more accurate at some values than others.

Method of Moments

MoM estimator

1. Compute moments $\mu_1, \mu_2, \dots$ in terms of parameters θ .
2. Invert expressions to find parameters in terms of moments.
3. Replace moments μ_k by sample moments $\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ to get MoM estimates for parameters.

Examples: For i.i.d. $X_1, \dots, X_n$

- ▶ $\text{Pois}(\lambda)$ has MoM estimator $\hat{\lambda} = \bar{X}$
- ▶ $\mathcal{N}(\mu, \sigma^2)$ has MoM estimators

$$\hat{\mu} = \bar{X} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Consistency of MoM

Instead of a general result, we will do this *ad-hoc*, using tools such as LLN, Continuous mapping, and Slutsky's.

Example: Gaussian model. The MoM estimators are

$$\hat{\mu} = \overline{X}_n, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \overline{X}_n^2.$$

The **LLN** estimates the population moments.

$$\overline{X}_n \xrightarrow{p} \mu, \quad \frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{p} \mathbb{E}[X_1^2] = \mu^2 + \sigma^2.$$

Continuous mapping then yields consistency.

$$\hat{\mu} \xrightarrow{p} \mu, \quad \hat{\sigma}^2 \xrightarrow{p} (\mu^2 + \sigma^2) - \mu^2 = \sigma^2.$$

Note: the empirical variance is **biased at finite n** , but **consistent**.

Example: Review 1, Q2

Let $X_1, \dots, X_n$ be i.i.d. from a shifted-exponential distribution:

$$f(x \mid \theta) = \begin{cases} e^{-(x-\theta)}, & x > \theta, \\ 0, & x \leq \theta, \end{cases} \quad \theta \in \mathbb{R}.$$

Equivalently, $X_i = Y_i + \theta$, where the Y_i are i.i.d. $\text{Expo}(1)$. Suppose we are given that $M_X(t) = e^{t\theta}/(1-t)$ for $t < 1$.

(a) Find the first moment $\mathbb{E}[X]$ using $M_X(t)$.

$$\mathbb{E}X = M'_X(0)$$

$$M'_X(t) = \frac{d}{dt} \frac{e^{t\theta}}{1-t} = \frac{\theta e^{t\theta}}{1-t} + \frac{e^{t\theta}}{(1-t)^2}$$

$$M'_X(0) = \left. \frac{d}{dt} \frac{e^{t\theta}}{1-t} \right|_{t=0} = \theta + 1$$

Example: Review 1, Q2

- (b) Find the method of moments estimator of θ in terms of $X_1, \dots, X_n$.
- (c) Is the estimator consistent? Justify your answer.

$$(b.) \quad \mu_1 = \theta + 1 \quad (\Rightarrow) \quad \theta = \mu_1 - 1$$

$$\hat{\theta}_{mom} = \bar{X} - 1$$

$$(c.) \quad \bar{X} \xrightarrow[\text{LLN}]{P} \theta + 1$$

By CLT

$$\hat{\theta}_{mom} = \bar{X} - 1 \xrightarrow{P} \theta, \text{ hence } \underline{\text{consistent}}.$$

