

# MATH357 - Statistics

## Summary

Winter 2025, Prof. Abbas Khalili.

Notes by Louis Meunier

1 Probability Prerequisites	1
2 Parametric Inference	2
3 Systematic Parameter Estimation	3
4 Confidence Intervals and Hypothesis Testing	4
5 Some MLEs and Such To Remember	4

## 1 Probability Prerequisites

**Definition 1:**  $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$  and  $S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$

**Theorem 1** (Properties of Normal Distributions): Let  $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ , then

- (i)  $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ ;
- (ii)  $\bar{X}_n$  and  $S_n^2$  are independent;
- (iii)  $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{(n-1)}^2$ ;
- (iv) If  $Z \sim \mathcal{N}(0, 1)$  and  $V \sim \chi_{(\nu)}^2$ ,  $\frac{Z}{\sqrt{V/\nu}} \sim t(\nu)$ . In particular,

$$\frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t(n-1).$$

- (v) If  $U \sim \chi_{(m)}^2$ ,  $V \sim \chi_{(n)}^2$  are independent rv's, then  $\frac{U/m}{V/n} \sim F(m, n)$ .

**Theorem 2** (Order Statistics): If  $X_1, \dots, X_n$  iid rv's with CDF  $F$ , the CDF's of the min, max order statistics are respectively

$$F_{X_{(1)}}(x) = 1 - [1 - F(x)]^n, \quad F_{X_{(n)}}(x) = [F(x)]^n,$$

and generally, for  $1 \leq j \leq n$ ,

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

**Theorem 3** (Convergence Theorems):

- (i) (Slutsky's) If  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{P} a$ , then  $X_n + Y_n \xrightarrow{d} X + a$ ,  $X_n Y_n \xrightarrow{d} aX$  and, if  $a \neq 0$ ,  $X_n/Y_n \xrightarrow{d} X/a$ .
- (ii) (Continuous Mapping Theorem) If  $X_n \xrightarrow{P, d} X$  and  $g$  continuous on a set  $C$  where  $P(X \in C) = 1$ , then  $g(X_n) \xrightarrow{P, d} g(X)$ .
- (iii) (WLLN) If  $X_i$  iid rv's with mean  $\mu$  and finite second moment,  $\bar{X}_n \xrightarrow{P} \mu$ .
- (iv) (First-Order Delta Method) If  $\sqrt{n}(X_n - \mu) \xrightarrow{d} V$  and  $g$  a function such that  $g'$  exist and is nonzero at  $x = \mu$ , then

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{d} g'(\mu) \cdot V.$$

- (v) (Second-Order Delta Method) If  $\sqrt{n}(X_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ , and  $g$  a function with  $g'(\mu) = 0$  but  $g''(\mu) \neq 0$ , then

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{d} \mathcal{N}(0, g'(\mu)^2 \sigma^2).$$

**Theorem 4** (Empirical CDF Properties): Let  $X_1, \dots, X_n$  be iid with cdf  $F$ . The ECDF is the rv defined by, for  $x \in \mathbb{R}$ ,  $F_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \leq x)$ . The following hold:

- (i)  $nF_n(x) \sim \text{Bin}(n, F(x))$ ; in particular,

$$\mathbb{E}[F_n(x)] = F(x), \quad \text{Var}(F_n(x)) = \frac{1}{n} F(x)(1 - F(x))$$

- (ii)  $\frac{\sqrt{n}(F_n(x) - F(x))}{\sqrt{F(x)(1 - F(x))}} \xrightarrow{d} \mathcal{N}(0, 1)$
- (iii)  $F_n(x) \rightarrow F(x)$

## 2 Parametric Inference

**Definition 2** (Qualities of Estimators):

- (i) The *bias* of an estimator  $\hat{\theta}$  of  $\theta$  is defined  $\text{Bias}(\hat{\theta}) = \mathbb{E}_\theta[\hat{\theta}] - \theta$ .  $\hat{\theta}$  is *unbiased* if it has zero bias.
- (ii) The *mean-squared error* (MSE) is defined  $\text{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$ .
- (iii) We say  $\hat{\theta}$  *unbiased* if  $\hat{\theta} \xrightarrow{P} \theta$

**Theorem 5** (Cramer-Rau Lower Bound): For a parametric family  $\{p(\cdot, \theta) : \theta \in \Theta\}$ , if  $T(\mathbf{X})$  an unbiased estimator of a function of a parameter  $\tau(\theta)$ , with finite variance, then

$$\text{Var}(T(\mathbf{X})) \geq \frac{[\tau'(\theta)]^2}{I(\theta)},$$

for every  $\theta \in \Theta$  in the, where  $I(\theta) := \mathbb{E}\left[\left(\frac{d}{d\theta} \log p_\theta(\mathbf{X})\right)^2\right]$  the *Fisher information* of the parametric family and assuming the denominator is finite, and moreover:

- (i)  $\{p_\theta : \theta \in \Theta\}$  has common support independent of  $\theta$
- (ii) for any  $x$  and  $\theta \in \Theta$ ,  $\frac{d}{d\theta} \log p_\theta(x) < \infty$
- (iii) for any statistic  $h(\mathbf{X})$  with finite first absolute moment, differentiation under the integral holds ie  $\frac{d}{d\theta} \int h(x)p(x) dx = \int h(x) \frac{d}{d\theta} p_\theta(x) dx$

Moreover, equality occurs iff there exists a function  $a(\theta)$  such that  $a(\theta)\{T(x) - \tau(\theta)\} = \frac{d}{d\theta} \log p(x; \theta)$ .

**Remark 1:** If  $p_\theta$  twice differentiable in  $\theta$  and  $\mathbb{E}\left[\frac{d}{d\theta} \log p_\theta(\mathbf{X})\right]$  differentiable “under the integral sign”, then  $I(\theta) = -\mathbb{E}\left[\frac{d^2}{d\theta^2} \log p_\theta(\mathbf{X})\right]$ .

If working with iid rv's, then the denominator becomes  $nI_1(\theta)$  where  $I_1(\theta)$  the Fisher information of a single rv.

**Theorem 6** (Neyman-Fisher Factorization): A statistic  $T(\mathbf{X})$ ,  $\mathbf{X} \sim p_\theta(\cdot)$  is called *sufficient* for  $\theta$  if the conditional distribution of  $\mathbf{X}$  given  $T(\mathbf{X}) = t$  is independent of  $\theta$ .  $T(\mathbf{X})$  is sufficient iff there are functions  $h(\cdot), g(\cdot; \theta)$  such that  $p_\theta(\mathbf{x}) = h(\mathbf{x})g(T(\mathbf{x}), \theta)$ .

**Theorem 7:** Any one-to-one function of a sufficient statistic is still sufficient.

**Theorem 8** (Minimal Sufficiency): A sufficient statistic is minimal if it is a function of every other sufficient statistic. For a parametric  $p_\theta(\cdot)$ , suppose  $T(\mathbf{x}) = T(\mathbf{y}) \Leftrightarrow \frac{p_\theta(\mathbf{x})}{p_\theta(\mathbf{y})}$  does not depend on  $\theta$ . Then,  $T(\mathbf{X})$  is minimally sufficient.

**Definition 3** (Completeness): An estimator  $\hat{\theta}$  is called *complete* if  $\mathbb{E}[g(\hat{\theta})] = 0$  for every  $\theta$  implies  $g = 0$  (a.s.).

**Theorem 9** (Rao-Blackwell): Let  $U(\mathbf{X})$  be unbiased for  $\tau(\theta)$  and  $T(\mathbf{X})$  sufficient, and define  $\delta(t) := \mathbb{E}_\theta[U(\mathbf{X}) \mid T(\mathbf{X}) = t]$ . Then  $\delta(\mathbf{X})$  is unbiased for  $\tau(\theta)$ , and has smaller variance than  $U(\mathbf{X})$ .

**Theorem 10** (Lehmann-Scheffé): Let  $T(\mathbf{X})$  be complete and sufficient and  $U(\mathbf{X}) = h(T(\mathbf{X}))$  unbiased with finite second moment, then  $U(\mathbf{X})$  is the UMVUE for  $\tau(\theta)$ .

**Remark 2:** Combine these two theorems to systematically construct UMVUEs starting from an (arbitrary) unbiased estimator and a complete and sufficient statistic.

### 3 Systematic Parameter Estimation

**Definition 4** (Method of Moments): The *method of moments* estimator(s) for rv's  $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} f_\theta$  is given by solving the system

$$\frac{1}{n} \sum_{i=1}^n X_i^j = \mu_j(\theta) := \mathbb{E}[X_i^j],$$

for  $j$  as high as we need for the system of equations to have solutions.

**Definition 5** (Minimum Likelihood Estimation (MLE)): An estimator  $\hat{\theta}_n$  is said to be an MLE of a parametric family if it maximizes the likelihood (resp. log likelihood) function (for any post-experimental data  $\mathbf{x}$ )

$$L_n : \Theta \rightarrow [0, \infty) \quad \left( \begin{array}{l} \ell_n : \Theta \rightarrow (-\infty, \infty) \\ \ell_n(\theta) = \log L_n(\theta) \end{array} \right).$$

If differentiable, one can solve for the (at least a candidate) MLE by solving the likelihood equations  $\partial_\theta L_n = 0$  or equivalently  $\partial_\theta \ell_n = 0$ .

*Remark 3:* Since log monotonic increasing, the likelihood/log-likelihood functions are equivalent and thus one should use which ever one is more convenient (lots of parametric families have exponentials, so using log is helpful).

**Theorem 11** (Properties of MLEs): We assume ["the regularity conditions"](#).

- (i) (Invariance) If  $\hat{\theta}$  the MLE of  $\theta$  and  $\tau(\theta)$  a function of  $\theta$ , then  $\tau(\hat{\theta})$  the MLE of  $\tau(\theta)$ .
- (ii)  $\hat{\theta}$  is consistent.
- (iii)  $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, [I_1^{-1}(\theta_0)])$  where  $\theta_0$  the "true value".
- (iv) (1st Bartlett Identity)  $\mathbb{E}_\theta \left[ \frac{\partial \log f(X)}{\partial \theta} \right] = 0$ .

**Definition 6** (Bayesian Estimation): Let  $\mathbf{X} \sim p_\theta$  where  $\theta$  also random, with pdf/pmf  $\pi(\theta)$ , called the *prior distribution* of  $\theta$ . The *posterior distribution* is defined as  $\pi(\theta|\mathbf{x})$ , which by Baye's is proportional to  $p_\theta(\mathbf{x})\pi(\theta)$ . A *loss function*  $L(\delta(\mathbf{X}), \theta)$  is a function assigning a "penalty" to an estimator  $\delta(\mathbf{X})$ , for instance the  $L^2$ -loss given by  $(\delta(\mathbf{X}) - \theta)^2$ . Baye's risk given a loss function  $L$  is defined

$$R(\delta) := \mathbb{E}_\pi \left[ \mathbb{E}_{\mathbf{X}|\theta} [L(\delta(\mathbf{X}), \theta)] \right].$$

Then, Baye's estimator is simply  $\hat{\delta}(\mathbf{X}) := \operatorname{argmin}_\delta R(\delta)$ .

**Theorem 12:** For  $L$  the  $L^2$ -loss function, the Baye's estimator is

$$\hat{\delta}(\mathbf{X}) = \mathbb{E}_{\theta|\mathbf{X}=\mathbf{x}}[\theta|\mathbf{X}].$$

*Remark 4:* So, given  $p_\theta$  and  $\pi(\theta)$ , the typical steps to finding  $\hat{\delta}(\mathbf{X})$  are:

- (i) compute  $p_\theta(\mathbf{x})\pi(\theta)$ , and deduce the distribution of  $(\theta|\mathbf{X})$ ;
- (ii) hopefully the distribution found in (i) has a well-known mean, which is then equal to the Baye's estimator  $\hat{\delta}(\mathbf{X})$  by the previous theorem.

## 4 Confidence Intervals and Hypothesis Testing

**Theorem 13** (Neyman-Pearson Lemma): Let

$$\phi(\mathbf{X}) := \begin{cases} 1 & \text{if } p(\mathbf{X}; \theta_1) > k \cdot p(\mathbf{X}; \theta_0) \\ 0 & \text{if } p(\mathbf{X}; \theta_1) < k \cdot p(\mathbf{X}; \theta_0) \end{cases},$$

and either if equal, where  $k$  is such that  $P_{\mathcal{H}_0}(\text{rejecting } \mathcal{H}_0) = \alpha$ . Then,  $\phi$  is the UMP test in the class of all tests at significance level  $\alpha$ .

## 5 Some MLEs and Such To Remember

Distribution	Sufficient Statistic	UMVUE	ME	MLE
Exponential, $f(x, \theta) = h(x)c(\theta) \exp(\omega(\theta)T_1(x))$	$\sum_{i=1}^n T_1(X_i)$	$\frac{1}{n} \sum_{i=1}^n T_1(X_i)$		
Poisson( $\lambda$ )	$f(\sum_{i=1}^n X_i)$	$\bar{X}_n$		$\bar{X}_n$
$\mathcal{U}(0, \theta)$		$\frac{n+1}{n} X_{(n)}$		
$\mathcal{N}(\mu, \sigma^2)$ $\mu, \sigma^2$ unknown	$(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$			
Ber( $\theta$ )	$\sum_{i=1}^n X_i$			

--	--	--	--	--