Chapter 26

Review for the Final Exam

Lecture 26 | Devore 9e: Ch. 2–6

The final is cumulative, with the weight on Lectures 18–25 (joint distributions, CLT, estimation). This chapter is the review class: a checklist, a formula sheet, and the worked problems.

26.1 Checklist

Topic	**You should be able to**
Bayes	invert a conditional; recognize the base-rate trap
Transformations	get f_Y from f_X via the cdf
Named distributions	recognize the scenario and quote pdf, mean, variance
Joint distributions	normalize, marginalize, condition; *draw the region*
Covariance/correlation	compute from a joint pmf/pdf; interpret the sign
CLT	approximate $\bar{X}$; know when it applies
Estimation	check unbiasedness; method of moments; MLE
Confidence intervals	build $\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$; say what 95% means

26.2 Master formula sheet

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \qquad \mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(A_i)\mathbb{P}(B \mid A_i)}{\sum_j \mathbb{P}(A_j)\mathbb{P}(B \mid A_j)}$$

$$\mathbb{E}[h(X)] = \int h(x) f(x)\,\mathrm{d}x, \qquad \mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

$$f_Y(y) = f_X(g^{-1}(y))\left|\tfrac{d}{dy} g^{-1}(y)\right|, \qquad f_X(x) = \int f(x,y)\,\mathrm{d}y$$

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)}, \qquad \mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mu_X \mu_Y$$

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}, \qquad \mathrm{Var}\Big(\sum a_i X_i\Big) = \sum a_i^2 \sigma_i^2 \text{ (independent)}$$

$$\mathbb{E}[\bar{X}] = \mu, \quad \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}, \qquad \bar{X} \overset{\text{approx}}{\sim} \mathrm{Normal}\Big(\mu, \frac{\sigma^2}{n}\Big)$$

$$\mathbb{E}[S^2] = \sigma^2, \qquad \ell(\theta) = \sum_i \log f(x_i; \theta)$$

$$\mathbb{E}[X^k] = \overline{X^k} \text{ (method of moments)}, \qquad \mu \in \bar{x} \pm z_{\alpha/2} \tfrac{\sigma}{\sqrt{n}} \quad (\sigma \text{ known})$$

26.3 Worked problems

Problem 1 (Bayes)

A disease affects 3% of the population. A test satisfies $\mathbb{P}(+ \mid D) = 0.92$ and $\mathbb{P}(- \mid D^c) = 0.88$. *(a)* Find $\mathbb{P}(+)$. *(b)* Find $\mathbb{P}(D \mid +)$.

Solution. $\mathbb{P}(D) = 0.03$, $\mathbb{P}(D^c) = 0.97$, and $\mathbb{P}(+ \mid D^c) = 1 - 0.88 = 0.12$.

$$\mathbb{P}(+) = \mathbb{P}(+ \mid D)\mathbb{P}(D) + \mathbb{P}(+ \mid D^c)\mathbb{P}(D^c) = (0.92)(0.03) + (0.12)(0.97) = 0.0276 + 0.1164 = 0.144.$$

$$\mathbb{P}(D \mid +) = \frac{\mathbb{P}(+ \mid D)\mathbb{P}(D)}{\mathbb{P}(+)} = \frac{0.0276}{0.144} = 0.1917.$$

Again: a positive result raises the probability from 3% only to 19%. □

Problem 2 (transformation)

$X \sim \mathrm{Normal}(0,1)$ and $Y = e^X$. Find $\mathbb{P}\{Y > e^{1.5}\}$.

Solution. exp is increasing, so the event is unchanged by taking logs:

$$\mathbb{P}\{Y > e^{1.5}\} = \mathbb{P}\{e^X > e^{1.5}\} = \mathbb{P}\{X > 1.5\} = 1 - \Phi(1.5) = 0.0668.$$

No density calculation is needed — that is the point of working with the cdf. □

Problem 3 (Weibull)

$X \sim \text{Weibull}(\alpha = 3, \beta = 10)$, so $F(x) = 1 - e^{-(x/10)^3}$ for $x > 0$. *(a)* Find the median $t_{0.5}$. *(b)* Find $\mathbb{P}\{5 \le X \le 15\}$.

Solution. (a) Solve $F(t_{0.5}) = 0.5$:

$$1 - e^{-(t_{0.5}/10)^3} = 0.5 \;\Rightarrow\; e^{-(t_{0.5}/10)^3} = \tfrac{1}{2} \;\Rightarrow\; \left(\tfrac{t_{0.5}}{10}\right)^3 = \ln 2 \;\Rightarrow\; t_{0.5} = 10(\ln 2)^{1/3} \approx 8.85.$$

(b) $\mathbb{P}\{5 \le X \le 15\} = F(15) - F(5) = e^{-(5/10)^3} - e^{-(15/10)^3} = e^{-1/8} - e^{-27/8} \approx 0.8825 - 0.0342 = 0.8483.$ □

Problem 4 (beta)

$X \sim \text{Beta}(3,3)$, so $f(x) = 30x^2(1-x)^2$ on $(0,1)$. *(a)* Find $\mathbb{E}[X]$ and $\text{Var}(X)$. *(b)* Find $\mathbb{P}\{0.4 \le X \le 0.6\}$.

Solution. (a) By the formulas for $\text{Beta}(\alpha, \beta)$,

$$\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta} = \frac{3}{6} = \frac{1}{2}, \qquad \text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{9}{36 \cdot 7} = \frac{1}{28}.$$

(b) $\displaystyle\int_{0.4}^{0.6} 30x^2(1-x)^2\,\mathrm{d}x = 30\Big[\tfrac{x^3}{3} - \tfrac{x^4}{2} + \tfrac{x^5}{5}\Big]_{0.4}^{0.6} = 0.6826 - 0.3174 = 0.3651.$ □

Problem 5 (joint pdf)

$f_{X,Y}(x,y) = c(x+y)$ for $0 < x < 1$, $0 < y < 1$. *(a)* Find c. *(b)* Find f_X.

Solution. (a)

$$1 = \int_0^1\!\int_0^1 c(x+y)\,\mathrm{d}x\,\mathrm{d}y = c\int_0^1 \Big(\tfrac{1}{2} + y\Big)\,\mathrm{d}y = c\Big(\tfrac{1}{2} + \tfrac{1}{2}\Big) = c \qquad \Longrightarrow \qquad c = 1.$$

(b) $\displaystyle f_X(x) = \int_0^1 (x+y)\,\mathrm{d}y = x + \tfrac{1}{2}$ for $0 < x < 1$.

The density is symmetric in x and y, so the same computation gives $f_Y(y) = y + \frac{1}{2}$ on $(0,1)$. Note $f_X(x)f_Y(y) = (x+\frac{1}{2})(y+\frac{1}{2}) \ne x+y$, so X and Y are dependent. □

Problem 6 (covariance and correlation from a table)

(X,Y) is uniform on the six pairs $(0,1),(1,1),(0,2),(2,2),(1,3),(2,3)$, each with probability $\frac{1}{6}$. Find $\mathbb{E}[X]$, $\mathbb{E}[Y]$, $\text{Cov}(X,Y)$, $\rho_{X,Y}$.

Solution. Marginals: each of $X = 0,1,2$ has probability $\frac{2}{6} = \frac{1}{3}$, and likewise each of $Y = 1,2,3$. So

$$\mathbb{E}[X] = \tfrac{1}{3}(0+1+2) = 1, \qquad \mathbb{E}[Y] = \tfrac{1}{3}(1+2+3) = 2.$$

$$\mathbb{E}[XY] = \tfrac{1}{6}\big(0\cdot 1 + 1\cdot 1 + 0\cdot 2 + 2\cdot 2 + 1\cdot 3 + 2\cdot 3\big) = \tfrac{14}{6} = \tfrac{7}{3},$$

$$\text{Cov}(X,Y) = \tfrac{7}{3} - (1)(2) = \tfrac{1}{3}.$$

$$\mathbb{E}[X^2] = \tfrac{1}{3}(0+1+4) = \tfrac{5}{3} \;\Rightarrow\; \text{Var}(X) = \tfrac{5}{3} - 1 = \tfrac{2}{3}, \qquad \mathbb{E}[Y^2] = \tfrac{1}{3}(1+4+9) = \tfrac{14}{3} \;\Rightarrow\; \text{Var}(Y) = \tfrac{14}{3} - 4 = \tfrac{2}{3}.$$

$$\rho_{X,Y} = \frac{1/3}{\sqrt{(2/3)(2/3)}} = \frac{1/3}{2/3} = \frac{1}{2}.$$

□

Problem 7 (MLE)

$X_1, \ldots, X_n$ i.i.d. Exponential(λ), $f(x;\lambda) = \lambda e^{-\lambda x}$. *(a)* Write the likelihood. *(b)* Find $\hat{\lambda}_{\text{MLE}}$. *(c)* Is it unbiased?

Solution. (a) $L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n \exp\Big(-\lambda \sum_{i=1}^{n} x_i\Big)$.

(b) $\ell(\lambda) = n \log \lambda - \lambda \sum_i x_i$, so $\ell'(\lambda) = \frac{n}{\lambda} - \sum_i x_i = 0$ gives

$$\hat{\lambda}_{\text{MLE}} = \frac{n}{\sum_i X_i} = \frac{1}{\bar{X}}.$$

(c) No. We know $\mathbb{E}[\bar{X}] = 1/\lambda$, but $\mathbb{E}\big[1/\bar{X}\big] \neq 1/\mathbb{E}[\bar{X}]$ — the reciprocal is a non-linear function. Concretely, $\sum_i X_i \sim \text{Gamma}(n, 1/\lambda)$ by gamma additivity, so for $n \geq 2$

$$\mathbb{E}\Big[\frac{1}{\sum_i X_i}\Big] = \int_0^\infty \frac{1}{s} \cdot \frac{\lambda^n s^{n-1} e^{-\lambda s}}{\Gamma(n)}\, \mathrm{d}s = \frac{\lambda\, \Gamma(n-1)}{\Gamma(n)} = \frac{\lambda}{n-1},$$

and therefore $\mathbb{E}[\hat{\lambda}] = n \cdot \frac{\lambda}{n-1} = \frac{n}{n-1}\lambda > \lambda$: the MLE overshoots on average. It is, however, asymptotically unbiased, and $\frac{n-1}{n}\hat{\lambda}$ is exactly unbiased. □

Remark 26.1 (Final advice)**.** Three habits that recover the most points:

1. **Draw the region** before setting up any double integral.
2. **Go through the cdf** for anything involving a transformation.
3. **Sanity-check the answer**: probabilities lie in $[0, 1]$, variances are non-negative, $|\rho| \leq 1$, and a mean should land inside the support.