COSE 382 HW 7 Solutions

---

1. A fair 6-sided die is rolled once. Find the expected number of additional rolls needed to obtain a value at least as large as that of the first roll.

Solution:

Let X be the result of the first roll and Y be the number of rolls to obtain a value $\geq X$. The conditional distribution of Y given $X = k$ is $FS\left(\frac{(7-k)}{6}\right)$, i.e.,

$$P(Y \mid X = h) \sim FS\left(\frac{7-k}{6}\right)$$

$$\begin{aligned} E(Y) &= \sum_{k=1}^{6} E(Y \mid X = k) P(X = k) = \sum_{k=1}^{6} \frac{b}{7-k} \cdot \frac{1}{6} \\ &= \sum_{k=1}^{6} \frac{1}{k} \end{aligned}$$

2. Let X_1, X_2 be i.i.d., and let $\bar{X} = \frac{1}{2}(X_1 + X_2)$ be the sample mean. In many statistics problems, it is useful or important to obtain a conditional expectation given $\bar{X}$. As an example of this, find $E(w_1X_1 + w_2X_2|\bar{X})$, where w_1, w_2 are constants with $w_1 + w_2 = 1$.

Solution:

$$2E\left(\omega_1 X_1 + \omega_2 X_2 \mid \bar{X}\right) = \omega_1 E\left(X_1 \mid \bar{X}\right) + \omega_2 E\left(X_2 \mid \bar{X}\right)$$

Note that $E\left(X_1 \mid \bar{X}\right) = E\left(X_2 \mid \bar{X}\right)$ by symmetry and

$$E\left(X_1 + X_2 \mid \bar{X}\right) = \left(X_1 + X_2\right) E(1|\bar{X}) = \left(X_1 + X_2\right)$$

Hence, $E\left(X_1 \mid \bar{X}\right) = E\left(X_2 \mid \bar{X}\right) = \bar{X}$ Thus,

$$E\left(w_1 X_1 + w_2 X_2 \mid \bar{X}\right) = w_1 \bar{X} + w_2 \bar{X} = \bar{X}$$

3. Let X be the height of a randomly chosen adult man, and Y be his father's height, where X and Y have been standardized to have mean 0 and standard deviation 1. Suppose that (X, Y) is Bivariate Normal, with $X, Y \sim N(0, 1)$ and $\text{Corr}(X, Y) = \rho$.

(a) Find a constant c (in terms of ρ) and an r.v. V such that $Y = cX + V$, with V independent of X.

(b) Find a constant d (in terms of ρ) and an r.v. W such that $X = dY + W$, with W independent of Y.

(c) Find $E(Y|X)$ and $E(X|Y)$.

Solution:

a) Let $V := Y - cX$. To V and X be independent $\operatorname{Corr}(V, X) = 0$ (since V and X me also bivariate)

$$\begin{aligned} \operatorname{cor}(Y, X) &= E((Y - cX)X) \\ &= E(YX) - c\left(X^2\right) = 0 \\ \Rightarrow c &= \rho \end{aligned}$$

b) By the same argument, $d = \rho$!

c)

$$\begin{aligned} E(Y \mid X) &= E(\rho X + V \mid X) = \rho E(X \mid X) + E(V \mid X) \\ &= \rho X \\ E(X \mid Y) &= E(\rho Y + W \mid X) = \rho E(Y \mid Y) + E(W|Y) \\ &= \rho Y \end{aligned}$$

Solution:

4. Let X and Y be random variables with finite variances, and let $W = Y - E(Y|X)$. This is a residual: the difference between the true value of Y and the predicted value of Y based on X.

(a) Compute $E(W)$ and $E(W|X)$.

(b) Compute $Var(W)$, for the case that $W|X \sim N(0, X^2)$ with $X \sim N(0, 1)$.

Solution:

a)

$$\begin{aligned} E(W) &= E(Y) - E(E(Y \mid X)) = EY - EY = 0 \\ E(W \mid X) &= E(Y \mid X) - E(E(Y \mid X) \mid X) \\ &= E(Y \mid X) - E(Y \mid X) = 0 \end{aligned}$$

b)

$$\begin{aligned} &\operatorname{Var}\ (W) = \operatorname{Var}(E(W \mid X)) + E(\operatorname{Var}(W \mid X)) \\ &= \operatorname{Var}(0) + E\left(X^2\right) = 1 \end{aligned}$$

5. Show that if $E(Y|X) = c$ is a constant, then X and Y are uncorrelated.

Solution:
When $E(Y \mid X) = C$,

$$\begin{aligned} E(Y) &= E(c) = c \\ E(XY) &= E(E(XY \mid X)) = E(XE(Y \mid X)) \\ &= E(cX) = cE(X) = E(Y)E(X), \end{aligned}$$

6. In a national survey, a random sample of people are chosen and asked whether they support a certain policy. Assume that everyone in the population is equally likely to be surveyed at each step, and that the sampling is with replacement. Let n be the sample size, and let $\hat{p}$ and p be the proportion of people who support the policy in the sample and in the entire population, respectively. Show that for every $c > 0$,

$$P(|\hat{p} - p| > c) \leq \frac{1}{4nc^2}$$

Solution:
Let $X \sim \text{Bin}(n, p)$, then $\hat{p} = \frac{X}{n}$

$$\begin{aligned}
E(\hat{p}) &= \frac{E(X)}{n} = \frac{np}{n} = p \\
\text{Var}(\hat{p}) &= \frac{\text{Var}(X)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}
\end{aligned}$$

By Chebyshev inequality

$$p(|\hat{p} - p| > c) \leq \frac{\text{Var}(\hat{p})}{c^2} = \frac{p(1-p)}{nc^2}$$

Since $p(1-p) \leq \frac{1}{4}$ for $\forall p \in (0, 1)$

$$P(|\hat{p} - p| > c) \leq \frac{1}{4nc^2}$$

7. For i.i.d. r.v.s $X_1, \cdots, X_n$ with mean μ and variance σ^2, find a value of n which will ensure that there is at least a 99% chance that the sample mean will be within 2 standard deviations of the true mean μ.

Solution:
We have to find n such that

$$P\left(\left|\bar{X}_n - \mu\right| > 2\sigma\right) \leq 0.01$$

By Chebyshev inequality.

$$P\left(\left|\bar{X}_n - \mu\right| > 2\sigma\right) \leq \frac{\text{Var}(\bar{X}_n)}{(2\sigma)^2} = \frac{\frac{\sigma^2}{n}}{4\sigma^2} = \frac{1}{4n}$$

8. Let X and Y be i.i.d. positive r.v.s, and let $c > 0$. For each part below, fill in the appropriate equality or inequality symbol. If no relation holds in general, write "?".

(a) $E(X) \qquad \sqrt{E(X^2)}$

(b) $P(X > c) \qquad E(X^3)/c^3$

(c) $E(X^3) \qquad \sqrt{E(X^2)E(X^4)}$

(d) $P(|X + Y| > 2) \qquad \frac{1}{10}E((X + Y)^4)$

(e) $E(Y|X) \qquad E(Y|X + 3)$

(f) $P(X+Y>2) \qquad (EX+EY)/2$

Solution:

a) $E(X)$ v.s. $\sqrt{E\left(X^2\right)}$ equivalently $E^2(X)$ v.s. $E\left(X^2\right)$.
Since $\operatorname{Var}(X)=E\left(X^2\right)-E^2(X)\geq 0$,

$$E\left(X^2\right)\geq E^2(X)\Rightarrow\sqrt{E\left(X^2\right)}\geq E(X)$$

b)

$$P(X>c)=P\left(X^3>c^3\right)=P\left(\left|X^3\right|>c^3\right)$$

By Markov inequality,

$$P\left(X^3>c^3\right)\leq\frac{E\left(X^3\right)}{c^3}$$

(c)

$$E\left(X^3\right)=E\left(X^2\cdot X\right)=\left|E\left(X^2\cdot X\right)\right|$$

By Cauchy-Schwartz

$$\begin{aligned}\left|E\left(X^2\cdot X\right)\right|&\leq\sqrt{E\left(X^4\right)\cdot E\left(X^2\right)}\\ E\left(X^3\right)&\leq\sqrt{E\left(X^4\right)\cdot E\left(X^2\right)}\end{aligned}$$

d) $P(|X+Y|>2)=P\left((X+Y)^4\geq 16\right)$ By Markov inequality

$$P\left((X+Y)^4\geq 16\right)\leq\frac{E\left((X+Y)^4\right)}{16}\leq\frac{1}{10}E\left((X+Y)^4\right)$$

e) $E(Y\mid X)=E(Y\mid X+3)$

f)

$$\begin{aligned}P(X+Y>2)=&\quad p(|X+Y|\geq 2)\\ P(|X+Y|\geq 2)\leq\frac{E(X+Y)}{2}&=\frac{E(X)+E(Y)}{2}\end{aligned}$$

9. Consider i.i.d. Pois(λ) r.v.s $X_1,X_2,\cdots$. The MGF of X_j is $M(t)=e^{\lambda(e^t-1)}$.

(a) Find the MGF $M_n(t)$ of the sample mean $\bar{X}_n=\frac{1}{n}\sum_{j=1}^n X_j$

(b) Find the limit of $M_n(t)$ as $n\to\infty$

Solution:

a)

$$\begin{aligned}M_n(t)=E\left(e^{t\bar{X}_n}\right)&=E\left(e^{\frac{t}{n}(X_1+\cdots+X_n)}\right)\\ &=\left(E\left(e^{\frac{t}{n}X_1}\right)\right)^n=e^{n\lambda\left(e^{t/n}-1\right)}\end{aligned}$$

b)

$$\lim_{n\to\infty}M_n(t)=e^{t\lambda}\text{ since }\bar{X}_n\to\lambda\text{ as }n\to\infty$$

10. Let $Y = e^X$, with $X \sim \text{Expo}(3)$.

(a) Find the mean and variance of Y .

(b) For $Y_1, \cdots, Y_n$ i.i.d. with the same distribution as Y, what is the approximate distribution of the sample mean $\bar{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$ when n is large?

Solution:

a)

$$E(Y) = \int_0^\infty e^x \left(3e^{-3x}\right) dx = \frac{3}{2}$$

$$E\left(Y^2\right) = \int_0^\infty e^{2x} \left(3e^{-3x}\right) dx = 3$$

$$E(Y) = \frac{3}{2}, \text{Var}(Y) = 3 - \frac{9}{4} = 3/4$$

b) By Central Limit Theorem, $Y_n \simeq N\left(\frac{3}{2}, \frac{3}{4n}\right)$ for large n