MATH 7241: Problem Set #3

Due date: Friday October 7

SOLUTIONS

Reading: relevant background material for these problems can be found in the class notes, and in Ross (Chapters 2,3,5) and in Grinstead and Snell (Chapters 1,2 3, 6).

$\boxed{\textbf{Exercise 1}}$ Let $N, X_1, X_2, \ldots$ be independent random variables, where the X_k are identically distributed with $\mathbb{E}[X_k] = \mu$ and $\mathrm{VAR}[X_k] = \sigma^2$. Also $\mathrm{RAN}(N) = \{1, 2, 3, \ldots\}$, and both mean and variance of N are finite. Show that

$$\mathrm{VAR}\left(\sum_{k=1}^{N} X_k\right) = \sigma^2\,\mathbb{E}[N] + \mu^2\,\mathrm{VAR}[N]$$

[Hint: let $Y = \sum_{k=1}^{N} X_k$, and note that $\mathrm{VAR}[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$. Use conditioning on N to compute $\mathbb{E}[Y^2]$, and the use the result from class about $\mathbb{E}[Y]$].

From class we have

$$\mathbb{E}[Y] = \mu\,\mathbb{E}[N].$$

Now

$$\mathbb{E}[Y^2 \mid N = n] = \mathbb{E}\left[\left(\sum_{k=1}^{n} X_k\right)^2 \,\Big|\, N = n\right]$$

$$= \mathbb{E}\left[\left(\sum_{k=1}^{n} X_k\right)^2\right] \quad \text{(b/c indep.)}$$

$$= \mathrm{VAR}\left[\sum_{k=1}^{n} X_k\right] + \left(\mathbb{E}\left[\sum_{k=1}^{n} X_k\right]\right)^2$$

$$= n\sigma^2 + n^2\mu^2.$$

$$\Rightarrow \quad \mathbb{E}[Y^2 \mid N] = N\sigma^2 + N^2\mu^2$$

$$\Rightarrow \quad \mathbb{E}[Y^2] = \mathbb{E}\left[\mathbb{E}[Y^2 \mid N]\right]$$

$$= \mathbb{E}[N]\sigma^2 + \mathbb{E}[N^2]\mu^2.$$

$$= \mathbb{E}[N]\sigma^2 + VAR[N]\mu^2 + \mathbb{E}[N]^2\mu^2.$$

$$\Rightarrow \quad VAR[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$$

$$= \mathbb{E}[N]\sigma^2 + VAR[N]\mu^2.$$

Exercise 2 Suppose X has a uniform distribution on $[0, 2]$, and Y is uniformly distributed on $[0, X^2]$. Find the expected value of XY.

Condition on X:

$$\mathbb{E}[XY \mid X = x] = x \underbrace{\mathbb{E}[Y \mid X = x]}_{Y \sim U[0,\, x^2]}$$

$\Rightarrow$ mean $= \frac{x^2}{2}$

$$= x\left(\frac{x^2}{2}\right)$$

$$= \frac{x^3}{2}.$$

$$\Rightarrow \mathbb{E}[XY \mid X] = \tfrac{1}{2} X^3.$$

$$\Rightarrow \mathbb{E}[XY] = \mathbb{E}\big[\,\mathbb{E}[XY \mid X]\,\big]$$

$$= \mathbb{E}\left[\tfrac{1}{2} X^3\right]$$

pdf of X

$$= \frac{1}{2} \int_0^2 x^3 \cdot \left(\frac{1}{2}\right) dx$$

$$= \frac{1}{4} \cdot \frac{x^4}{4} \Bigg|_0^2$$

$$= 1.$$

Exercise 3 Let X be an exponential random variable with rate λ, and let $t > 0$ be a fixed number.

a) Use the memoryless property to compute

$$\mathbb{E}[X \mid X > t]$$

b) By combining the result of part (a) with the total probability probability formula for $\mathbb{E}[X]$, compute

$$\mathbb{E}[X \mid X \leq t]$$

a) Given $X > t \Rightarrow X = t + X'$ where X' is exp. rate λ.

$$\Rightarrow \mathbb{E}[X \mid X > t] = \mathbb{E}[t + X' \mid X > t]$$

$$= t + \mathbb{E}[X' \mid X > t]$$

$$= t + \frac{1}{\lambda}.$$

b) Total probability:

$$\mathbb{E}[X] = \mathbb{E}[X \mid X > t]\, \mathbb{P}(X > t) + \mathbb{E}[X \mid X \leq t]\, \mathbb{P}(X \leq t)$$

$$\Rightarrow \frac{1}{\lambda} = \left(t + \frac{1}{\lambda}\right) e^{-\lambda t} + \mathbb{E}[X \mid X \leq t]\left(1 - e^{-\lambda t}\right).$$

$$\Rightarrow \mathbb{E}[X \mid X \leq t] = \frac{\frac{1}{\lambda} - \frac{1}{\lambda} e^{-\lambda t} - t e^{-\lambda t}}{1 - e^{-\lambda t}}$$

$$= \frac{1}{\lambda} - t \frac{e^{-\lambda t}}{1-e^{-\lambda t}}$$

Exercise 4 Let X_n be the symmetric random walk starting at 0. In class we derived the formula

$$P(X_n = n - 2k) = \frac{n!}{(n-k)!\,k!}\,2^{-n}, \quad k = 0, 1, 2, \ldots, n.$$

Define $x = k/n$ and use Stirling's formula $n! \sim n^n \sqrt{2\pi n} e^{-n}$ to show that when n and k are both large, the following asymptotic formula holds:

$$P(X_n = n - 2k) \simeq \frac{1}{\sqrt{2\pi n x(1-x)}}\, e^{-n[\log 2 - h(x)]}$$

where

$$h(x) = -x \log x - (1-x) \log(1-x)$$

Use Stirling's Formula:

$$\frac{n!}{(n-k)!\,k!} \approx \frac{n^n \sqrt{2\pi n}\, e^{-n}}{(n-k)^{n-k} \sqrt{2\pi(n-k)}\, e^{-(n-k)}\; k^k \sqrt{2\pi k}\, e^{-k}}$$

$$= \frac{1}{\sqrt{2\pi}} \frac{n^n}{(n-k)^{n-k} k^k} \cdot \sqrt{\frac{n}{(n-k)\,k}}$$

Write $x = \frac{k}{n} \Rightarrow k = nx$

$$\Rightarrow \text{RHS} = \frac{1}{\sqrt{2\pi}} \frac{n^n}{[n(1-x)]^{n-nx} (nx)^{nx}} \frac{1}{\sqrt{n(1-x)x}}$$

$$= (2\pi n(1-x)x)^{-\frac{1}{2}} \left[(1-x)^{n(1-x)} x^{nx}\right]^{-1}$$

$$= (2\pi n(1-x)x)^{-\frac{1}{2}} \exp\left[-n(1-x)\log(1-x) - nx\log x\right]$$

$$= (2\pi n x(1-x))^{-\frac{1}{2}} \exp\left[n\, h(x)\right]$$

$$\Rightarrow \mathbb{P}(X_n = n-2k) = \frac{n!}{(n-k)!\,k!} 2^{-n}$$

$$\simeq \left[2\pi n\, x(1-x)\right]^{-\frac{1}{2}}.$$

$$\times \exp\left[n\, h(x) - n \log 2\right].$$

$\boxed{\textbf{Exercise 5}}$ Suppose that $\{X_i\}$ are IID uniform random variables on the interval $[-1, 1]$. Let Z be a standard normal random variable. Using the CLT, find the number a so that

$$\lim_{n\to\infty} P\Big(\sum_{i=1}^{n} X_i \geq \sqrt{n}\Big) = P(Z \geq a)$$

[Hint: you will need to find the mean and variance of X, which is uniform on $[-1, 1]$].

$X \sim U[-1,1] \Rightarrow$ pdf $f(x) = \begin{cases} \frac{1}{2} & -1 \leq x \leq 1 \\ 0 & \text{else} \end{cases}$

$\Rightarrow \mathbb{E}[X] = 0$ (easy!)

$\mathbb{E}[X^2] = \int_{-1}^{1} x^2 \cdot \frac{1}{2}\, dx = \int_0^1 x^2\, dx = \frac{1}{3}.$

$\Rightarrow \mathrm{VAR}[X] = \frac{1}{3}.$

Let $Z_n = \dfrac{\sum_{i=1}^{n} X_i - n(0)}{\sqrt{n}\sqrt{\frac{1}{3}}}$.

Then CLT says $Z_n \to Z \sim N(0,1)$ as $n \to \infty$.

Now

$\mathbb{P}\Big(\sum_{i=1}^{n} X_i \geq \sqrt{n}\Big)$

$= \mathbb{P}\big(Z_n \geq \sqrt{3}\big)$

$$\to \mathbb{P}(Z > \sqrt{3}) \quad \text{as } n \to \infty$$

$$\Rightarrow \quad a = \sqrt{3}.$$

Exercise 6 A random variable can take values $\{1, 2, 3, 4\}$. The observed frequencies from 200 measurements are $\{85, 70, 25, 20\}$ respectively. The null hypothesis is

$$H_0 : p_1 = 0.4, \ p_2 = 0.3, \ p_3 = 0.2, \ p_4 = 0.1$$

and the alternative hypothesis is that these are not the probabilities. Use goodness of fit and the chi-square distribution to test H_0 at the 1% significance level: find the expected frequencies, find the number of degrees of freedom, find the critical value from the tables, state the decision rule, find the test statistic of the data, and state your conclusion.

Outcomes	1	2	3	4
Probs.	0.4	0.3	0.2	0.1
Expected Frequencies	80	60	40	20
Observed Frequencies	85	70	25	20

H_0

$N = 200$

Number of degrees of freedom : $df = 4 - 1 = 3.$

1% significance level : $\alpha = 0.01. = 1 - p$

Critical value (from tables) $\chi^2_{0.99,3} = 11.345.$

Let D = goodness of fit test statistic

Decision rule: reject H_0 if $D > 11.345$

Compute D:

$$D = \frac{(80-85)^2}{80} + \frac{(60-70)^2}{60} + \frac{(40-25)^2}{40} + \frac{(20-20)^2}{20}$$

$$= \frac{25}{80} + \frac{100}{60} + \frac{225}{40}$$

$$= \frac{5}{16} + \frac{5}{3} + \frac{45}{8}$$

$$= \frac{95}{16} + \frac{5}{3}$$

$$= 7.604$$

<u>Conclusion</u>: since $D = 7.604 < 11.345$

$\Rightarrow$ we do not reject H_0.

The data supports the null hypothesis.