PROBABILITY: QUIZ 2

MARK DITSWORTH

1. Problem 1

Indicate True of False for each statment.

1.1. Part A.

X and Y are independent random variables. X is uniformly distributed on $[-2,2]$. Y is uniformly distributed on $[-1,5]$. If $Z = X + Y$, then $f_Z(3) = 1/6$.

True.
This PDF of the sum of two random variables is found through convolution, where $f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\ dx$.

$f_Z(3) = \int_{-\infty}^{\infty} f_X(x) f_Y(3-x)\ dx$
$f_X(x)$ in non-zero only for $x \in [-2,2]$. $f_Y(y)$ is only non-zero for $y \in [-1,5]$, thus $f_Y(3-x)$ is only non-zero for $x \in [-2,4]$. The limits of integration is the union of these two sets: $[-2,2]$.
$f_Z(3) = \int_{-2}^{2} f_X(x) f_Y(3-x)\ dx = \left(\frac{1}{6}\right)\left(\frac{1}{4}\right)(2+2) = \frac{1}{6}$

1.2. Part B.

If X is a Gaussian random variable with mean 0 and variance 1, then the density function of of $Z = |X|$ is equal to $2f_X(x), z \geq 0$.

True. Since X is symmetric about 0, $Z = |X|$ means the probability of $-x$ adds to the probability of x, causing the $f_Z(z) = 2f_X(z)$ where $z \geq 0$.

1.3. Part C.

The sum of a random number (N) of independent Gaussian random variables with zero mean and unit variance results in a Gaussian random variable, regardless of the distribution of N.

False
For $Y = X_1 + X_2 + \cdots + X_N$, the transform of Y can be found as such.
$M_Y(s) = M_N(s)|_{e^s = M_X(s)}$. Let N be Gaussian.

$M_N(s) = e^{s^2/2}$ and since X is Gaussian, $M_X(s) = e^{s^2/2}$.

Thus, $M_Y(s) = \sqrt{(e^s)^s}|_{e^s = M_X(s)} = \sqrt{(e^{s^2/2})^s}$, which is clearly not Gaussian.

1.4. **Part D.** If X and Y are independent random variables, both exponentially distributed with parameters λ_1 and λ_2 respectively, then the random variable $Z = \min\{X, Y\}$ is also exponentially distributed.

True
The CDF $F_Z(z) = P(Z \leq z) = P(\min\{X,Y\} \leq z)$ is not immediately evident, but $1 - F_Z(z)$ is.

$$\begin{gathered}
1 - F_Z(z) = P(Z > z) = P(\min\{X,Y\} > z) \\
= P(X > z, Y > z) = P(X > z)\ P(Y > z) \\
= (1 - F_X(z))\,(1 - F_Y(z)) = \left(1 - 1 + e^{-\lambda_1 z}\right)\left(1 - 1 + e^{-\lambda_2 z}\right) \\
1 - F_Z(z) = e^{-(\lambda_1+\lambda_2)z} \\
f_Z(z) = (\lambda_1 + \lambda_2)e^{-(\lambda_1+\lambda_2)z}
\end{gathered}$$

1.5. **Part E.** Let the transform associated with random variable X be $M_X(s) = \left(\frac{e^s}{1-s}\right)^{15}$. $\mathbf{E}[X]$ is equal to 30.

True

$$\begin{gathered}
\mathbf{E}[X] = \frac{d}{ds}M_X(s)|_{s=0} = 15\left(\frac{e^s}{1-s}\right)^{14}\frac{(1-s)e^s + e^s}{(1-s)^2}|_{s=0} \\
\mathbf{E}[X] = 15\left(\frac{1}{1-0}\right)^{14}\frac{(1)(1)+1}{(1-0)^2} = 15\ (2) = 30
\end{gathered}$$

1.6. **Part F.** X and Y are independent random variables. Y is normal with mean 0 and variance 1, X is uniform on $[0, 1]$. $Z = X + Y$. The conditional density of Z given X, $f_{Z|X}(z|x)$ is normal with mean x and variance 1.

True
Z given some x is the normal random variable Y plus x, $Z|x = Y + x$. Thus, $\mathbf{E}[Z|X] = \mathbf{E}[Y] + x$ and $\mathrm{Var}(Z|X) = \mathrm{Var}(Y)$.

1.7. **Part G.** Var(Z)=2.

False
Since X and Y are independent, $\mathrm{Var}(X+Y)$=$\mathrm{Var}(X)$+$\mathrm{Var}(Y)$
So, $\mathrm{Var}(Z) = \frac{1}{12} + 1 \neq 2$.

1.8. **Part H.** $\mathbf{E}[X|Z = -1] = -1$

False
X is bounded by $[0, 1]$, so it cannot take the value of -1.

1.9. **Part I.** Cov(X,Z) = Var(X)

True

$$\text{Cov}(X,Z) = \mathbf{E}[XZ] - \mathbf{E}[X]\mathbf{E}[Z]$$

$$= \mathbf{E}[X(X+Y)] - \mathbf{E}[X]\mathbf{E}[X+Y]$$

$$= \mathbf{E}[X^2] + \mathbf{E}[XY] - \mathbf{E}[X]^2 - \mathbf{E}[X]\mathbf{E}[Y]$$

$$= \mathbf{E}[X^2] - \mathbf{E}[X]^2 + \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$$

$$= \text{Var}(X) + \text{Cov}(X,Y)$$

And since X and Y are independent, Cov(X,Y)=0, thus Cov(X,Z) = Var(X)

1.10. **Part J.** $Z = \mathbf{E}[X|Z] + \mathbf{E}[Y|Z]$

True
$\mathbf{E}[Z|Z] = \mathbf{E}[X+Y|Z]$ and since conditional expectation is linear, $\mathbf{E}[Z|Z] = Z = \mathbf{E}[X|Z] + \mathbf{E}[Y|Z]$

2. Problem 2

2.1. **Part A.** Find the least squares estimate of Y given $X = x$, for all possible values of x.

$$\mathbf{E}[Y|X] = \begin{cases} \frac{1}{2} & 0 \leq X < 1 \\ X - \frac{1}{2} & 1 \leq X \leq 2 \end{cases}$$

2.2. **Part B.** Let $g(x)$ be the estimate from Part A. Find $\mathbf{E}[g(X)]$ and $\mathrm{Var}(g(X))$.

$$\mathbf{E}[g(X)] = \int_0^2 g(x) f_X(x)\ dx = \frac{1}{2}\left(\int_0^1 \frac{1}{2}\ dx + \int_1^2 x - \frac{1}{2}\ dx\right)$$

$$\mathbf{E}[g(X)] = \frac{1}{2}\left(\frac{1}{2} + 1 - 0\right) = \frac{3}{4}$$

$$\mathbf{E}[g(X)^2] = \int_0^2 g(x)^2 f_X(x)\ dx = \frac{1}{2}\left(\frac{1}{4} + \int_1^2 (x - \frac{1}{2})^2\ dx\right) = \frac{1}{2}(\frac{1}{4} + \frac{13}{12}) = \frac{2}{3}$$

$$\mathrm{Var}(g(X)) = \mathbf{E}[g(X)^2] - \mathbf{E}[g(X)]^2 = \frac{2}{3} - \frac{9}{16} = \frac{5}{48} \approx 0.1042$$

2.3. **Part C.** Find the mean squared error $\mathbf{E}[(Y - g(X))^2]$. Is it the same as $\mathbf{E}[\mathrm{var}(Y|X)]$?

Since $g(X) = \mathbf{E}[Y|X]$,

$$\mathbf{E}[(Y - g(X))^2] = \mathbf{E}[(Y - \mathbf{E}[Y|X])^2] = \mathbf{E}[\mathbf{E}[(Y - \mathbf{E}[Y|X])^2|X]]$$

$$= \mathbf{E}[\mathbf{E}[(Y|X - \mathbf{E}[Y|X])^2]] = \mathbf{E}[\mathrm{Var}(Y|X)]$$

Since Y is uniform for all X, $\mathrm{Var}(Y|X) = \frac{1}{12}(Y_1 + 1 - Y_1)^2 = \frac{1}{12}$, $0 \leq X \leq 2$
$\mathbf{E}[\mathrm{Var}(Y|X)] = \frac{1}{2}\int_0^2 \frac{1}{12}\ dx = \frac{1}{12}$

3. Problem 3

Each year, an editor is sent a random number of books for review. The number of books received can be modeled as a Poisson random variable N with mean μ. Each book contains a random number of typos, modeled by a Poisson random variable with mean λ. Let B_i denote the number of typos in book i. Assume all random variables are independent. The editor finds typos with probability p, independent of all other findings, and other random variables.

There are two different payment options:

Option 1. \$1 for each typo found

Option 2. \$1 for each book where at least 1 typo is found

Let X_i be the amount of money the editor receives for book i and T the total amount of money the editor receives in a year.

3.1. Part A. The the PMF of X_i under option 1.

The PMF of X_i under option 1 is the PMF of the number of typos in book i. For each typo is book i, the editor has a p probability of catching it. Thus, the PMF is a binomial distribution with the number of trials the random variable of B_i.

$$\mathbf{P}(X_i = x | B_i = b) = \binom{b}{x} p^x (1-p)^{b-x}$$

Since X_i and B_i are independent,

$$\mathbf{P}(X_i = x) = \sum_{b=x}^{\infty} \mathbf{P}(X_i = x | B_i = b)\mathbf{P}(B_i = b)$$

$$\mathbf{P}(X_i = x) = \sum_{b=x}^{\infty} \binom{b}{x} p^x (1-p)^{b-x} \frac{\lambda e^{-\lambda}}{b!}$$

$$\mathbf{P}(X_i = x) = \frac{(p\lambda)^x}{x!} e^{-\lambda} \sum_{b=0}^{\infty} \frac{(\lambda(1-p))^b}{b!} = \frac{(p\lambda)^x}{x!} e^{-p\lambda}$$

This is a Poisson PMF with mean $p\lambda$.

3.2. Part B. Find $M_T(s)$ under option 1.

$T = X_1 + X_2 + \ldots X_N$. Thus, $M_T(s) = M_N(s)|_{e^s = M_X(s)}$

$$M_N(s) = e^{\mu(e^s - 1)} \qquad M_X(s) = e^{p\lambda(e^s - 1)}$$

$$M_T(s) = e^{\mu\left(e^{p\lambda(e^s-1)} - 1\right)}$$

3.3. **Part C.** Find $\mathbf{P}(T = 2)$ under option 2.

Since T is a discrete random variable, we can find probabilities through differentiation of $M_T(s)$ as such:

$$\mathbf{P}(T = t) = \frac{1}{t!}\frac{d^t}{d(e^s)^t}M(s)|_{e^s=0}$$

$$\mathbf{P}(T = 2) = \frac{1}{2}\frac{d^2}{d(e^s)^2}M_T(s)|_{e^s=0}$$

$$= \left[\frac{\mu p\lambda}{2}\exp\{\mu e^{p\lambda(e^s-1)-\mu+p\lambda e^s-p\lambda}\left(\mu p\lambda e^{p\lambda(e^s-1)} + p\lambda\right)\right]|_{e^s=0}$$

$$\mathbf{P}(T = 2) = \frac{1}{2}\mu(p\lambda)^2 e^{-p\lambda}e^{\mu(e^{-p\lambda-1})}\left(\mu e^{-p\lambda} + 1\right)$$

3.4. **Part D.** Find $\mathbf{E}[T]$.

$$\mathbf{E}[T] = \frac{d}{ds}M_T(s)|_{s=0} = e^{\mu\left(e^{p\lambda(e^s-1)}-1\right)}\mu e^{p\lambda(e^s-1)}p\lambda e^s|_(s = 0)$$

$$\mathbf{E}[T] = \mu p\lambda$$

3.5. **Part E.** Find $\mathrm{Var}(T)$.

$$\mathrm{Var}(T) = \mathbf{E}[T^2] - \mathbf{E}[T]^2$$

$$\mathbf{E}[T^2] = \frac{d^2}{ds^2}M_T(s)|_{s=0} = (\mu p\lambda)^2 + \mu(p\lambda)^2 + \mu p\lambda$$

$$\mathrm{Var}(T) = (\mu p\lambda)^2 + \mu(p\lambda)^2 + \mu p\lambda - (\mu p\lambda)^2 = \mu p\lambda(p\lambda + 1)$$

3.6. **Part F.** Find the PMF of X_i under option 2.

Under option 2, $X_i = 1$ if at least 1 typo is found in book i, and 0 otherwise (Binomial distribution). From Part A, we see that the number of found typos is book i is a Poisson random variable with mean $p\lambda$. Thus, the probability that $X_i = 0$ is $e^{-p\lambda}$. The full PMF is:

$$X_i = \begin{cases} e^{-p\lambda} & x = 0 \\ 1 - e^{-p\lambda} & x = 1 \\ 0 & \text{otherwise} \end{cases}$$

3.7. **Part G.** Find $\mathbf{E}[T]$ under option 2.

Each book viewed by the editor, there is a ρ probability that he or she will get a dollar for the book. There are N number of attempts ate getting dollars. Thus, T has a binomial distribution with the number of trials being a Poisson random variable. Following a similar procedure as Part A:

$$\mathbf{P}(T = t) = \sum_{n=t}^{N} \binom{n}{t} \rho^t (1-\rho)^{n-t} \frac{\mu^n}{n!} e^{-\mu} = \frac{(\mu\rho)^t}{t!} e^{-\mu\rho}$$

This is a Poisson distribution with mean $\mu\rho$. From Part G, we saw that the probability ρ of receiving a dollar for book i is $1 - e^{-p\lambda}$. Therefore, we have:

$$\mathbf{E}[T] = \mu\rho = \mu(1 - e^{-p\lambda})$$

3.8. **Part H.** Which option should the editor choose?

We will answer this by finding the option with the highest expected value for T. Under option 1, $\mathbf{E}[T] = \mu p \lambda$. Under option 2, $\mathbf{E}[T] = \mu(1 - e^{-p\lambda})$.

Both μ and λ are out of the editor's control. But p is determined by the editor's skill in spotting typos. As p increases towards 1, the expected yearly payout under option 1 grows linearly towards $\mu\lambda$. Whereas under option 2 it is bounded above by $\mu(1 - e^{-\lambda})$. Note that $1 - e^{-\lambda}$ grows sub-linearly, so $\mu\lambda$ will always be greater than $\mu(1 - e^{-\lambda})$.

Therefore, option 1 will result in the highest expected value of T.