Probability

Solutions to Exam B, Fall 2015

Solution Q1:

(A) We will use the following identity (Pascal's triangle): For $1 \le k \le n$,

$$\binom{n}{k-1} + \binom{n}{k} = \frac{n!}{(k-1)!(n-k)!} \cdot \left(\tfrac{1}{n+1-k} + \tfrac{1}{k}\right)$$
$$= \frac{n!}{(k-1)!(n-k)!} \cdot \frac{n+1}{k(n+1-k)} = \binom{n+1}{k}.$$

Now, for any $k \in \{1, \ldots, n+1\}$ we have using the law of total probability and the independence of X, Y,

$$\mathbb{P}[Z = k] = \mathbb{P}[X + Y = k] = \mathbb{P}[X = k-1, Y = 1] + \mathbb{P}[X = k, Y = 0]$$
$$= \mathbb{P}[X = k-1] \cdot p + \mathbb{P}[X = k] \cdot (1-p)$$
$$= \binom{n}{k-1} p^{k-1}(1-p)^{n-k+1} \cdot p + \binom{n}{k} p^k (1-p)^{n-k} \cdot (1-p)$$
$$= \binom{n+1}{k} p^k (1-p)^{n+1-k}.$$

For $k = 0$ we have

$$\mathbb{P}[Z = 0] = \mathbb{P}[X = 0, Y = 0] = \mathbb{P}[X = 0] \cdot \mathbb{P}[Y = 0] = (1-p)^{n+1}.$$

since these all add up to 1, we get that $R_Z = \{0, 1, \ldots, n+1\}$ and for $k \in \{0, 1, \ldots, n+1\}$,

$$f_Z(k) = \binom{n+1}{k} p^k (1-p)^{n+1-k},$$

which is exactly the $\text{Bin}(n+1, p)$ density.

(B) This we prove by induction on m. For $m = 1$, since $\text{Bin}(1, p) = \text{Ber}(p)$, we get that $A + B \sim \text{Bin}(n+1, p)$ by the previous item.

For the induction step, assume the claim holds for m and we will prove it for $m+1$. So $A \sim \mathrm{Bin}(n,p)$ and $B \sim \mathrm{Bin}(m+1,p)$.

Let $C \sim \mathrm{Bin}(m,p)$ and $D \sim \mathrm{Ber}(p)$ be such that A,B,C,D are independent. Note that by the previous item, B and $C+D$ have the same distribution. Thus, it suffices to prove that $A+C+D \sim \mathrm{Bin}(n+m+1,p)$.

By induction, $A+C \sim \mathrm{Bin}(n+m,p)$ because A,C are independent.

However, since $A+C$ and D are independent, by the previous item again, $(A+C)+D \sim \mathrm{Bin}(n+m+1,p)$, which completes the induction.

(C) Let $A \sim \mathrm{Bin}(n,\frac{1}{2})$ and $B \sim \mathrm{Bin}(m,\frac{1}{2})$ be independent, for $m \geq n$. Then, by the law of total probability, because $m \geq n$,

$$\begin{aligned}\mathbb{P}[A+B=n] &= \sum_{k=0}^{n} \mathbb{P}[A=k, B=n-k] = \sum_{k=0}^{n} \mathbb{P}[A=k] \cdot \mathbb{P}[B=n-k] \\ &= \sum_{k=0}^{n} \binom{n}{k} 2^{-n} \cdot \binom{m}{n-k} 2^{-m} \\ &= 2^{-(n+m)} \cdot \sum_{k=0}^{n} \binom{n}{k} \cdot \binom{m}{n-k}.\end{aligned}$$

But by the previous item, $A+B \sim \mathrm{Bin}(n+m,\frac{1}{2})$, so

$$\mathbb{P}[A+B=n] = 2^{-(n+m)} \cdot \binom{n+m}{n}.$$

Solution Q2:

(A) Let R be the range of X and Y (they have the same range, since they have the same distribution). Then, by the law of total probability and independence,

$$\begin{aligned}\mathbb{P}[X < Y] &= \sum_{r \in R} \mathbb{P}[X < Y, X = r] = \sum_{r \in R} f_X(r) \cdot \mathbb{P}[Y > r] \\ &= \sum_{r \in R} f_X(r) \cdot (1 - F_Y(r)) = \sum_{r \in R} f_Y(r) \cdot (1 - F_X(r)) \\ &= \sum_{r \in R} \mathbb{P}[Y < X, Y = r] = \mathbb{P}[Y < X],\end{aligned}$$

where we have used the fact that $f_X = f_Y$ and $F_X = F_Y$. This is the first equality.

Now, $\Omega = \{Y < X\} \uplus \{X < Y\} \uplus \{X = Y\}$, so

$$1 = \mathbb{P}[X < Y] + \mathbb{P}[Y < X] + \mathbb{P}[X = Y] = 2\,\mathbb{P}[X < Y] + p.$$

A simple rearrangement completes the proof.

(B) Note that $\mathbb{P}[X < Y] = \mathbb{P}[(X, Y) \in A]$. So, simply,

$$\mathbb{P}[X < Y] = \int\int_A f_{X,Y}(x, y) dy dx = \int_{-\infty}^{\infty} \int_x^{\infty} f_{X,Y}(x, y) dy dx.$$

Now, X, Y are independent and have the same distribution. So $f_{X,Y}(x, y) = f_X(x) f_Y(y)$. But then,

$$\mathbb{P}[X < Y] = \int_{-\infty}^{\infty} f_X(x) \int_x^{\infty} f_Y(y) dy dx = \int_{-\infty}^{\infty} f_X(x)(1 - F_Y(x)) dx.$$

We can do the same calculation with $f_{Y,X}$. Again, $f_{Y,X}(y, x) = f_Y(y) f_X(x)$ by independence. So

$$\mathbb{P}[Y < X] = \int_{-\infty}^{\infty} f_Y(y)(1 - F_X(y)) dy,$$

just as above.

We now use the fact that $f_X = f_Y$ and $F_X = F_Y$ to get that

$$\mathbb{P}[Y < X] = \int_{-\infty}^{\infty} f_Y(y)(1 - F_X(y)) dy = \int_{-\infty}^{\infty} f_X(y)(1 - F_Y(y)) dy = \mathbb{P}[X < Y].$$

Finally, since $\Omega = \{Y < X\} \uplus \{X < Y\} \uplus \{X = Y\}$, and since X, Y are continuous, and thus $\mathbb{P}[X = Y] = 0$, we get that

$$1 = \mathbb{P}[X < Y] + \mathbb{P}[Y < X] = 2\,\mathbb{P}[X < Y].$$

(C) The main observation is that if we define $A_i = \{X_i > X_{i-1}\}$ for all $i \geq 2$, we get

$$N = \sum_{i=2}^{n} \mathbf{1}_{\{X_i > X_{i-1}\}} = \sum_{i=2}^{n} \mathbf{1}_{A_i}.$$

In the previous item we showed that $\mathbb{P}[A_i] = \frac{1}{2}$ for all i. By linearity we immediately get

$$\mathbb{E}[N] = \sum_{i=2}^{n} \mathbb{P}[A_i] = \frac{n-1}{2}.$$

As for the variance, we first compute $\mathrm{Cov}(\mathbf{1}_{A_i}, \mathbf{1}_{A_j})$ for $j > i$:

If $j > i + 1$ then since $A_j \in \sigma(X_j, X_{j-1})$ and $A_i \in \sigma(X_i, X_{i-1})$, we get that A_j, A_i are independent, which is to say that

$$\mathbb{E}[\mathbf{1}_{A_j}\mathbf{1}_{A_i}] = \mathbb{P}[A_j \cap A_i] = \mathbb{P}[A_j] \cdot \mathbb{P}[A_i] = \mathbb{E}[\mathbf{1}_{A_j}] \cdot \mathbb{E}[\mathbf{1}_{A_i}].$$

That is, $\mathrm{Cov}(\mathbf{1}_{A_i}, \mathbf{1}_{A_j}) = 0$ for all $j > i + 1$.

If $j = i + 1$ then

$$\mathbb{E}[\mathbf{1}_{A_j}\mathbf{1}_{A_i}] = \mathbb{P}[A_j \cap A_i] = \mathbb{P}[X_{i+1} > X_i > X_{i-1}] = \tfrac{1}{6}.$$

So

$$\mathrm{Cov}(\mathbf{1}_{A_i}, \mathbf{1}_{A_{i+1}}) = \frac{1}{6} - \mathbb{E}[\mathbf{1}_{A_{i+1}}] \cdot \mathbb{E}[\mathbf{1}_{A_i}] = \frac{1}{6} - \frac{1}{4} = -\frac{1}{12}.$$

Now using the Pythagorean Theorem,

$$\begin{aligned}\operatorname{Var}[N] &= \sum_{i=2}^{n} \operatorname{Var}[\mathbf{1}_{A_i}] + 2\sum_{2\le i<j\le n} \operatorname{Cov}(\mathbf{1}_{A_i}, \mathbf{1}_{A_j}) \\ &= \frac{n-1}{4} + 2\sum_{i=2}^{n-1} \operatorname{Cov}(\mathbf{1}_{A_i}, \mathbf{1}_{A_{i+1}}) + 2\sum_{\substack{2\le i\le n \\ j>i+1}} \operatorname{Cov}(\mathbf{1}_{A_i}, \mathbf{1}_{A_j}) \\ &= \frac{n-1}{4} - 2\cdot\frac{n-2}{12} = n(\tfrac{1}{4} - \tfrac{1}{6}) - \tfrac{1}{4} + \tfrac{1}{3} = \frac{n+1}{12}.\end{aligned}$$

(We have used the fact that indicators are Bernoulli random variables, so $\operatorname{Var}[I] = \mathbb{E}[I](1-\mathbb{E}[I])$ for an indicator I.)

Solution Q3:

(A) First, set $Y = \frac{X}{\sigma}$. So $Y \sim \mathcal{N}(0,1)$. Then, since $\frac{\partial}{\partial y} e^{-y^2/2} = -ye^{-y^2/2}$,

$$\begin{aligned}I_1 &:= \int_0^\infty ye^{-y^2/2}dy = (-e^{-y^2/2})\big|_0^\infty = 1 \\ I_2 &:= \int_{-\infty}^0 ye^{-y^2/2}dy = (-e^{-y^2/2})\big|_{-\infty}^0 = -1.\end{aligned}$$

Thus,

$$\mathbb{E}[|Y|] = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^\infty |y|e^{-y^2/2}dy = \frac{1}{\sqrt{2\pi}}(I_1 - I_2) = \sqrt{\tfrac{2}{\pi}}.$$

Thus,

$$\mathbb{E}[|X|] = \sigma\,\mathbb{E}[|Y|] = \sqrt{\tfrac{2}{\pi}} \cdot \sigma.$$

(B) Since the total mass is 1,

$$\begin{aligned}1 &= \int\int f_{X,Y}(t,s)dsdt = \sqrt{\pi}c \cdot \int_0^1 \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\cdot(t/\sqrt{2})}|s|\exp(-\tfrac{s^2}{2(t/\sqrt{2})^2})dsdt \\ &= \sqrt{\pi}c \cdot \int_0^1 \mathbb{E}[|\mathcal{N}(0, t/\sqrt{2})|]dt = \sqrt{\pi}c \cdot \int_0^1 \sqrt{\tfrac{2}{\pi}} \cdot \tfrac{t}{\sqrt{2}}dt = c \cdot \tfrac{1}{2}.\end{aligned}$$

Thus $c = 2$.

(C) For $t \notin (0,1)$ we have $f_{X,Y}(t,s) = 0$ so $f_X(t) = 0$. For $0 < t < 1$ we get

$$f_X(t) = \int_{-\infty}^{\infty} f_{X,Y}(t,s)ds = 2\sqrt{\pi} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \cdot (t/\sqrt{2})} |s| \exp(-\tfrac{s^2}{2(t/\sqrt{2})^2}) ds$$
$$= 2\sqrt{\pi}\, \mathbb{E}[|\mathcal{N}(0, t/\sqrt{2})|] = 2\sqrt{\pi} \cdot \sqrt{\tfrac{2}{\pi}} \cdot \tfrac{t}{\sqrt{2}} = 2t.$$

That is,

$$f_X(t) = \begin{cases} 0 & t \notin (0,1) \\ 2t & t \in (0,1). \end{cases}$$

One may indeed check that this is a density.