Answers (Midterm 1 practice questions)

Online S520

1. (a) *I toss six fair coins. What is the probability exactly four of the coins show heads?*

   By counting or the binomial, this is $\binom{6}{4}/2^6 = 15/64$.

   (b) *I toss six fair coins. What is the probability that there are more heads than tails?*

   Let X be the number of heads. We need to find $P(X = 4) + P(X = 5) + P(X = 6)$. This is

$$\frac{\binom{6}{4} + \binom{6}{5} + \binom{6}{6}}{2^6} = \frac{22}{64}.$$

   (c) *I toss six fair coins. Given that there are more heads than tails, what is the conditional probability exactly four of the coins show heads?*

   From the definition of conditional probability, this is the answer to (a) divided by the answer to (b), which is 15/22.

   (d) *I toss six fair coins. What is the expected value and variance of number of coins that show heads?*

   Use the properties of the binomial: $EX = np = 6 \times 0.5 = 3$ and $\text{Var}(X) = np(1-p) = 6 \times 0.5 * 0.5 = 1.5$.

2. *Let X be a discrete random variable with probability mass function*

$$P(X = x) = \begin{cases} 0.2 & x = -1 \\ 0.3 & x = 2 \\ 0.3 & x = 4 \\ 0.2 & x = 7 \\ 0 & \text{otherwise.} \end{cases}$$

   (a) *Write down an expression for $F(y)$, the cumulative distribution function (CDF) of X, for all y-values from $-\infty$ to ∞.*

$$F(y) = P(X \le y) = \begin{cases} 0 & y < -1 \\ 0.2 & -1 \le y < 2 \\ 0.5 & 2 \le y < 4 \\ 0.8 & 4 \le y < 7 \\ 1 & y \ge 7 \end{cases}$$

   (b) *Find the expected value and variance of X.*

   $EX = 0.2 \times (-1) + 0.3 \times 2 + 0.3 \times 4 + 0.2 \times 7 = 3$.
   $EX^2 = 0.2 \times (-1)^2 + 0.3 \times 2^2 + 0.3 \times 4^2 + 0.2 \times 7^2 = 16$.
   $\text{Var}(X) = EX^2 - (EX)^2 = 16 - 3^2 = 7$.

   (c)

   (d) Let X_1 and X_2 be independent random variables with the same distribution as X. Compare the expected value and variance of $X_1 + X_2$ and $2X$.

$$\begin{aligned}
E(X_1 + X_2) &= E(X_1) + E(X_2) = 3 + 3 = 6 \\
Var(X_1 + X_2) &= Var(X_1) + Var(X_2) = 7 + 7 = 14 \\
E(2X) &= 2E(X) = 2 \times 3 = 6 \\
Var(2X) &= 2^2 Var(X) = 4 \times 7 = 28
\end{aligned}$$

$X_1 + X_2$ has the same expected value as $2X = X + X$, but smaller variance, because X_1 is independent of X_2.

(e) *Let X_1 and X_2 be independent random variables with the same distribution as X. What is the probability that $X_1 + X_2$ is positive?*

$X_1 + X_2$ is always positive unless both $X_1 = -1$ and $X_2 = -1$.

$$\begin{aligned}
P(X_1 = -1, X_2 = -1) &= 0.2 \times 0.2 = 0.04 \\
P(X_1 + X_2 > 0) &= 1 - 0.04 = 0.96.
\end{aligned}$$

3. *Let X be a continuous random variable with pdf*

$$f(x) = \begin{cases} 0.1 & 0 \le x < 2 \\ 0.2 & 2 \le x < 6 \\ 0 & \text{otherwise.} \end{cases}$$

(a) *Find F, the cumulative distribution function of X.*

$$F(y) = \begin{cases} 0 & y < 0 \\ 0.1y & 0 \le y < 2 \\ 0.2y - 0.2 & 2 \le y < 6 \\ 1 & y \ge 6 \end{cases}$$

(b) *Find the median of X.*

Setting $F(q_2) = 0.5$, we get a median of $q_2 = 3.5$.

(c) *Find the expected value of X.*

Taking the weighted average of the centers of the two blocks, this is $0.2 \times 1 + 0.8 \times 4 = 3.4$.

4. *Let X be a continuous random variable with probability density function (PDF)*

$$f(x) = \begin{cases} \frac{1}{2}x & 0 \le x < 1 \\ \frac{1}{2} & 1 \le x < 2 \\ \frac{1}{2}(3 - x) & 2 \le x < 3 \\ 0 & \text{otherwise.} \end{cases}$$

(a) $F(y)$ is a continuous function, so X is continuous. (Alternatively, $F(y)$ is differentiable at all points except $y = 1$ and $y = 4$, so a PDF exists.).

(b) *Find F, the cumulative distribution function of X.*

For $0 \le x \le 1$,

$$F(x) = \int_0^x \frac{1}{2} u \, du = \frac{1}{4} x^2.$$

For $1 \le x \le 2$,

$$\begin{aligned} F(x) &= P(X \le x) \\ &= P(X \le 1) + P(1 < X \le x) \\ &= F(1) + \int_1^x \frac{1}{2}\,du \\ &= \frac{1}{4} + \left(\frac{1}{2}x - \frac{1}{2}\right) \\ &= \frac{1}{2}x - \frac{1}{4}. \end{aligned}$$

For $2 \le x \le 3$,

$$F(x) = F(2) + \int_2^x \frac{1}{2}(3-x)\,du = -\frac{5}{4} + \frac{3x}{2} - \frac{1}{4}x^2.$$

This gives:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4}x^2 & 0 \le x < 1 \\ \frac{1}{2}x - \frac{1}{4} & 1 \le x < 2 \\ -\frac{5}{4} + \frac{3x}{2} - \frac{1}{4}x^2 & 2 \le x < 3 \\ 1 & x \ge 3 \end{cases}$$

(c) *Find the standard deviation of X. (Hint: Use the results above from calculus.)*

$$\begin{aligned} EX &= \int_0^1 \frac{1}{2}x^2\,dx + \int_1^2 \frac{1}{2}x\,dx + \int_2^3 \frac{1}{2}x(3-x)\,dx \\ &= \frac{1}{6}x^3\Big|_0^1 + \frac{1}{4}x^2\Big|_1^2 + \left(\frac{3x^2}{4} - \frac{x^3}{6}\right)\Big|_2^3 \\ &= \frac{1}{6} + 1 - \frac{1}{4} + \frac{9}{4} - \frac{5}{3} \\ &= \frac{3}{2}. \quad \text{(Not required)} \end{aligned}$$

$$
\begin{aligned}
EX^2 &= \int_0^1 \frac{1}{2}x^3\,dx + \int_1^2 \frac{1}{2}x^2\,dx + \int_2^3 \frac{1}{2}x^2(3-x)\,dx \\
&= \frac{1}{8}x^4\Big|_0^1 + \frac{1}{6}x^3\Big|_1^2 + \left(\frac{x^3}{2} - \frac{x^4}{8}\right)\Big|_2^3 \\
&= \frac{1}{8} + \frac{4}{3} - \frac{1}{6} + \frac{27}{8} - 2 \\
&= \frac{8}{3} \quad \text{(Not required)} \\
VarX &= \frac{8}{3} - \left(\frac{3}{2}\right)^2 \\
&= \frac{5}{12} \\
sd(X) &= \sqrt{\frac{5}{12}} \\
&\approx 0.645.
\end{aligned}
$$

5. *Let X be a standard normal random variable. Let $Y = |X|$.*

   (a) *Find $P(-1.5 < X < 2.5)$.*
   We can calculate this as the difference between the CDF of the standard normal at 2.5 and the CDF at -1.5:

   ```
   > pnorm(2.5) - pnorm(-1.5)
   [1] 0.9269831
   ```

   This gives about 93%.

   (b) *Find $P(Y > 1)$.*

$$
\begin{aligned}
P(Y > 1) &= P(|X| > 1) \\
&= P(X > 1 \text{ or } X < -1) \\
&= P(X > 1) + P(X < -1) \\
&= 1 - P(X \le 1) + P(X < -1)
\end{aligned}
$$

   = `(1 - pnorm(1)) + pnorm(-1)` = 31.7%.

   (c) *Find the 0.9-quantile of Y.*

   Suppose $y > 0$ is 0.9-quantile of Y

$$
\begin{aligned}
0.9 &= P(Y < y) \\
&= P(|X| < y) \\
&= P(-y < X < y) \\
0.95 &= P(X < y)
\end{aligned}
$$

   y is 0.95-quantile of X = `qnorm(.95)` = 1.645.

6. (a) *Let X be a normal random variable with mean -5 and standard deviation 10. Find $P(X > 0)$.*

   $P(X > 0) = 1 - P(X \leq 0) =$ `1 - pnorm(0, -5, 10)` $= 0.309$

   (b) *Let Z_1, Z_2, Z_3, Z_4, and Z_5 be independent standard normal random variables. What is the probability that at least two of the five variables are greater than 1?*

   The probability that one standard normal random variable is greater than 1 is: $P(Z > 1) =$ `1-pnorm(1)`. This is the success probability in each Bernoulli trial.

   Let A denote $\{$at least two of the five variables are greater than 1$\}$, then the complement A^c is $\{$at most one of the five variables is greater than 1$\}$, and $P(A) = 1 - P(A^c)$. $P(A^c)$ is the probability of observing no more than 1 success in this Binomial experiment with 5 trials and success probability `1-pnorm(1)`. So

   $$P(A^c) = \texttt{pbinom(1, 5, 1-pnorm(1))}$$

   $P(A) =$ `1 - pbinom(1, 5, 1-pnorm(1))` gives 0.181.

7. Let X, Y, and Z be independent standard normal random variables.

   (a) *Find $P(1 < X < 2)$.*

   We can calculate this as the difference between the CDF of the standard normal at 2 and the CDF at 1:

   `pnorm(2) - pnorm(1)` $= 0.136$.

   (b) *Find $P(1 < X^2 < 2)$.*

   $$\begin{aligned} P(1 < X^2 < 2) &= P(1 < X < \sqrt{2} \text{ or } -\sqrt{2} < X < -1) \\ &= P(1 < X < \sqrt{2}) + P(-\sqrt{2} < X < -1) \\ &= P(X < \sqrt{2}) - P(X < 1) + P(X < -1) - P(X < -\sqrt{2}) \end{aligned}$$

   =`pnorm(sqrt(2)) - pnorm(1) + pnorm(-1) - pnorm(-sqrt(2))` $= 0.160$.

   (c) *Find $P(1 < X + Y + Z < 2)$.*

   $X + Y + Z$ is also a normal random variable with mean 0 (0+0+0) and variance 3 (1+1+1), standard deviation $\sqrt{3}$.

   $P(1 < X + Y + Z < 2) = P(X + Y + Z < 2) - P(X + Y + Z < 1)$
   =`pnorm(2, sd=sqrt(3)) - pnorm(1, sd=sqrt(3))` $= 0.158$.

8. *I give a ten question true/false statistics test to a class of ten chimpanzees. The chimpanzees, who do not know any statistics, randomly guess true or false, independently for each question and independently of each other. A score of at least 8 out of 10 is required to pass the test. What is the probability that at least one of the chimpanzees passes the test?*

Two Binomial experiments are involved in this question. One is the score one chimpanzee can get from the 10-question test ($n_1 = 10, p_1 = 0.5$), and we want to know the probability that one chimpanzee can pass the test,
$p.pass = P(X \geq 8) = 1 - P(X \leq 7) = 1 - \texttt{pbinom(7, 10, 0.5)} = 0.0546875$.

The other Binomial experiment is the number of chimpanzees that pass the test ($n_2 = 10, p_2 = p.pass$). We want to know the probability that at least one chimpanzee passes the test,
$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \texttt{pbinom(0, 10, 0.0546875)}$

```
> prob.pass = 1 - pbinom(7, 10, 0.5)
> 1 - pbinom(0, 10, prob.pass)
[1] 0.4301586
```

There's a 43% chance at least one of the chimps passes.