

Stat 134: Conditional Probabilities, Distributions, & Expectations Review: Solutions

Brian Thorsen

May 2nd, 2018

1. Let $X_1 \sim \text{Geom}(p_1)$, $X_2 \sim \text{Geom}(p_2)$, $X_1 \perp X_2$, both on $\{1, 2, \dots\}$. Find:

- (a) $P(X_1 \leq X_2)$

$$P(X_1 \leq X_2) = \frac{p_1}{1 - q_1 q_2}$$

- (b) $P(X_1 = x \mid X_1 \leq X_2)$. Recognize $X_1 \mid X_1 \leq X_2$ as a named distribution, and state the parameter(s).

$$X_{1 \mid X_1 \leq X_2} \sim \text{Geom}(1 - q_1 q_2) \text{ on } \{1, 2, \dots\}$$

- (c) $P(X_1 = k \mid X_1 + X_2 = n)$ in the case $p_1 = p_2$. Recognize this as a named distribution and state the parameter(s).

$$X_{1 \mid X_1 + X_2 = n} \sim \text{Unif on } \{1, 2, \dots, n - 1\}$$

2. Let $Y \sim \text{Beta}(r, s)$. Conditioned on $Y = y$, let $X \sim \text{Geometric}(y)$ on $\{0, 1, 2, \dots\}$. For simplicity, assume $r, s > 1$.

- (a) What is $\mathbb{E}(X \mid Y = y)$?

$$\mathbb{E}(X \mid Y = y) = \frac{1 - y}{y}$$

- (b) Find $\mathbb{E}(X)$.

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}\left(\frac{1 - Y}{Y}\right) = \frac{s}{r - 1}$$

- (c) Find $P(X = x)$, for $x \in \{0, 1, 2, \dots\}$.

$$P(X = x) = \frac{\Gamma(r + s)\Gamma(r + 1)\Gamma(s + x)}{\Gamma(r)\Gamma(s)\Gamma(r + s + x + 1)}, \quad x \in \{0, 1, 2, \dots\}$$

3. Suppose a proportion p of a population has a gene m that makes them predisposed to migraines. Of these people, the number of migraines they experience in a year follows a Poisson process with rate λ_m per year, whereas the rest of the population experiences migraines according to a Poisson process with rate λ_x .

- (a) What is the probability that a randomly selected individual experiences no migraines in a given year?

$$pe^{-\lambda_m} + qe^{-\lambda_x}$$

- (b) Let N_t denote the number of migraines a randomly selected individual experiences in t years. Find $\mathbb{E}(N_t)$.

$$t(p\lambda_m + q\lambda_x)$$

- (c) Find $\text{Var}(N_t)$.

Let $\mu_t = \mathbb{E}(N_t) = t(p\lambda_m + q\lambda_x)$. Then,

$$\text{Var}(N_t) = p(t\lambda_m - \mu_t)^2 + q(t\lambda_x - \mu_t)^2 + \mu_t$$

Call this variance σ_t^2 ; we'll need it for part (e).

- (d) Given that a person experienced k migraines in a year, find the expected number of migraines they will have next year.

Let M be the event this individual has gene m . The idea here is that we condition on $N_{[0,1)} = k$ to update the likelihood that this individual has gene m , and then condition on this probability to find the expected number of migraines next year.

$$P(M|N_{[0,1)} = k) = \frac{pe^{\lambda_m} \frac{\lambda_m^k}{k!}}{pe^{\lambda_m} \frac{\lambda_m^k}{k!} + qe^{-\lambda_x} \frac{\lambda_x^k}{k!}}$$

Call this updated probability p_k . Then,

$$E(N_{[1,2)}|N_{[0,1)} = k) = p_k\lambda_m + q_k\lambda_x$$

- (e) Challenge: Find $\text{Corr}(N_{[0,1)}, N_{[1,2)})$, i.e. the correlation between number of migraines in consecutive years.

It may be tempting to say that because the consecutive intervals are independent, the correlation is zero. *This is not the case*, because as seen in part (d), the number of migraines in the first year updates the probability this individual has gene m . This is closely related to the Rule of Succession, as we have seen previously in lecture.

Computationally, this is a borderline intractable problem; we certainly wouldn't ask you to evaluate this on a final. You'll see why below.

Let $X = N_{[0,1)}$, $Y = N_{[1,2)}$. Then $\mathbb{E}(X) = \mathbb{E}(Y) = \mu_1$, and $\text{Var}(X) = \text{Var}(Y) = \sigma_1^2$. Using the definition of correlation,

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{SD(X)SD(Y)} = \frac{\mathbb{E}(XY) - \mu_1^2}{\sigma_1^2}$$

What remains is to evaluate $\mathbb{E}(XY)$. This is where the problem becomes intractable:

$$\begin{aligned} \mathbb{E}(XY) &= \mathbb{E}(\mathbb{E}(XY|Y)) \\ &= \mathbb{E}(X\mathbb{E}(Y|X)) \\ &= \mathbb{E}(X(p_X\lambda_m + q_X\lambda_x)) \end{aligned}$$

You can attempt this calculation at your own peril, but it is probably not the most productive use of your study time.