STAT 512: Statistical Inference	**Autumn 2022**

Quiz Session 5: Thinning Properties

Yikun Zhang *November 2, 2022*

Some parts of this notes are inspired by Professor Daniela Witten's talk with the title "Double dipping: problems and solutions, with application to single-cell RNA-sequencing data".

The Poisson distribution is interconnected with the binomial distribution from various aspects. One of the most well-known connections between these two types of discrete distributions is that the Poisson distribution can be derived as the limiting case of the binomial distribution as the number of trials goes to infinity while the expected number of successes remains fixed. See also Example 2.3.13 in Casella and Berger [2002] and Application 3.3.5 in Perlman [2020].

Problem 1 (Law of Rare Events)**.** *Let $X_n \sim$ Binomial(n, p_n), where $p_n = \frac{\lambda}{n}$ for some fixed $\lambda \in (0, \infty)$. Show that as $n \to \infty$, X_n converges to Poisson(λ) in distribution.*

Proof. Recall from Lecture 3 Notes that the moment generating function (MGF) of X_n is

$$\begin{aligned} M_{X_n}(t) &= \left[p_n e^t + (1 - p_n)\right]^n \\ &= \left[1 + \left(\frac{\lambda}{n}\right)(e^t - 1)\right]^n \\ &\to e^{\lambda(e^t - 1)} \end{aligned}$$

as $n \to \infty$ and for any $t \in (-\infty, \infty)$, where we use the fact that $\lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = e$ in the limit. As $M_{X_\infty} = e^{\lambda(e^t-1)}$ is the MGF of Poisson(λ), we conclude by Theorem 2.3.12 in Casella and Berger [2002] that X_n converges to Poisson(λ) in distribution. □

Remark 1. *One can also prove the result in Problem 1 by working directly with the probability mass function of X_n as:*

$$\begin{aligned} \mathbb{P}(X_n = k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k} \cdot \frac{\lambda^k}{k!}\left(1 - \frac{\lambda}{n}\right)^n \\ &\to \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

as $n \to \infty$.

Another connection between Poisson and binomial distributions is that a Poisson random variable conditional on its sum with another independent Poisson random variable will follow a binomial distribution. Recall from Homework 2 that if we have two random variables

$$X \sim \text{Poisson}(p\lambda), \quad Y \sim \text{Poisson}\left((1-p)\lambda\right)$$

with $p \in (0, 1)$ and X, Y are independent, then we know that $X + Y \sim$ Poisson(λ) and the conditional distribution of X given $X + Y = n$ satisfies that

$$X|X + Y = n \sim \text{Binomial}(n, p).$$

In Problem 2 below, we will prove that the converse of the above statement also holds, *i.e.*,

$$N \sim \text{Poisson}(\lambda) \text{ and } X|N = n \sim \text{Binomial}(n, p) \quad \Longrightarrow \quad X \sim \text{Poisson}(p\lambda) \text{ and } N - X \sim \text{Poisson}\left((1-p)\lambda\right).$$

This result is known as the *thinning property* of Poisson variables or processes; see Section 3.7.2 in Durrett [2019] or Section 5.1 in Stoyan et al. [2013].

Problem 2 (Poisson Thinning)**.** *Suppose that a random variable N follows the Poisson distribution with mean λ. Let $Y_1, Y_2, ...$ be an i.i.d. binary sequence with $\mathbb{P}(Y_i = 1) = p$. Consider $X = |\{m \le N : Y_m = 1\}| = \sum_{m=1}^{N} Y_m$.*

(a) Show that X follows a Poisson distribution with mean λp.

(b) Show that X and $N - X$ are independent.

(c) Calculate the (Pearson's) correlation $Cor(X, N)$.

More generally, we consider an i.i.d. sequence $Y_1, Y_2, ...$ with $\mathbb{P}(Y_i = j) = p_j$ for $j = 1, ..., k$, i.e., *Y_i follows a discrete distribution on $\{1, ..., k\}$. Let $N_j = |\{m \le N : Y_m = j\}|$.*

(d) Show that $N_1, ..., N_k$ are independent and N_j has a Poisson distribution with mean λp_j for $j = 1, ..., k$.

Proof. (a) There are multiple ways to prove the thinning property, among which we only present two different approaches.

Method A (Direct Approach): Notice that $X|N = n \sim \text{Binomial}(n, p)$. The probability mass function of X is given by

$$\begin{aligned}
\mathbb{P}(X = x) &= \sum_{n=x}^{\infty} \mathbb{P}\left(X = x, N = n\right) \quad \text{(Think about why the summation starts from } n = x\text{)} \\
&= \sum_{n=x}^{\infty} \mathbb{P}\left(X = x | N = n\right) \cdot \mathbb{P}(N = n) \\
&= \sum_{n=x}^{\infty} \binom{n}{x} p^x (1-p)^{n-x} \cdot \frac{\lambda^n}{n!} e^{-\lambda} \\
&= \sum_{n=x}^{\infty} \frac{n!}{(n-x)!x!} \cdot p^x \left[(1-p)\lambda\right]^{n-x} \cdot \frac{\lambda^x}{n!} \cdot e^{-\lambda(1-p)} \cdot e^{-\lambda p} \\
&= \frac{(\lambda p)^x}{x!} \cdot e^{-\lambda p} \sum_{n=x}^{\infty} \frac{\left[(1-p)\lambda\right]^{n-x}}{(n-x)!} \cdot e^{-\lambda(1-p)} \\
&= \frac{(\lambda p)^x}{x!} \cdot e^{-\lambda p} \sum_{y=0}^{\infty} \frac{\left[(1-p)\lambda\right]^{y}}{y!} \cdot e^{-\lambda(1-p)} \quad \text{let } y = n - x \\
&= \frac{(\lambda p)^x}{x!} \cdot e^{-\lambda p}
\end{aligned}$$

for any integer x, showing that the marginal distribution of X is Poisson(λp).

Method B (MGF Approach): By law of total expectation, the MGF of X is given by

$$\begin{aligned}
\mathbb{E} \exp(tX) &= \mathbb{E}\left[\mathbb{E}\left(e^{tX} | N\right)\right] \\
&= \mathbb{E}\left[(pe^t + 1 - p)^N\right] \quad \text{recall the MGF of Binomial}(N, p)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{n=0}^{\infty} \left(pe^t + 1 - p\right)^n \cdot \frac{\lambda^n}{n!} e^{-\lambda} \\
&= \exp\left[\lambda p e^t + \lambda(1-p)\right] \cdot e^{-\lambda} \\
&= \exp\left[\lambda p (e^t - 1)\right],
\end{aligned}
$$

which is the MGF of Poisson(λp). The result follows.

(b) By symmetry, one can easily show that $N - X$ follows a Poisson $(\lambda(1-p))$ distribution. Thus, for any integers x, y, we have that

$$
\begin{aligned}
\mathbb{P}(X = x, N - X = y) &= \mathbb{P}(X = x, N = x + y) \\
&= \mathbb{P}(X = x | N = x + y) \cdot \mathbb{P}(N = x + y) \\
&= \frac{(x+y)!}{x! \cdot y!} p^x (1-p)^y \cdot \frac{\lambda^{x+y}}{(x+y)!} \cdot e^{-\lambda} \\
&= \left[\frac{(\lambda p)^x}{x!} e^{-\lambda p}\right] \cdot \left[\frac{(\lambda(1-p))^y}{y!} e^{-\lambda(1-p)}\right] \\
&= \mathbb{P}(X = x) \cdot \mathbb{P}(N - X = y).
\end{aligned}
$$

Hence, X and $N - X$ are independent.

(c) By direct calculations, we have that

$$
\begin{aligned}
\mathrm{Cov}(X, N) &= \mathbb{E}(XN) - \mathbb{E}(X) \cdot \mathbb{E}(N) \\
&= \mathbb{E}\left[N \cdot \mathbb{E}(X|N)\right] - \lambda p \cdot \lambda \\
&= \mathbb{E}(N^2 p) - p\lambda^2 \\
&= p(\lambda + \lambda^2) - p\lambda^2 \\
&= p\lambda.
\end{aligned}
$$

Hence, the correlation between X and N is

$$
\begin{aligned}
\mathrm{Cor}(X, N) &= \frac{\mathrm{Cov}(X, N)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(N)}} \\
&= \frac{p\lambda}{\sqrt{p\lambda \cdot \lambda}} \\
&= \sqrt{p}.
\end{aligned}
$$

(d) Notice that $N_j | N = n \sim \text{Binomial}(n, p_j)$, because $N_j = \sum_{m=1}^{N} \mathbb{1}_{\{Y_m = j\}}$. One can follow the derivations in (a) to show that $N_j \sim \text{Poisson}(\lambda p_j)$ for $j = 1, ..., k$. To prove that $N_1, ..., N_k$ are independent, we again exploit the joint probability mass function of $(N_1, ..., N_k)$ and argue that

$$
\begin{aligned}
&\mathbb{P}(N_1 = n_1, ..., N_k = n_k) \\
&= \mathbb{P}\left(N_1 = n_1, ..., N_k = n_k \middle| N = \sum_{i=1}^{k} n_i\right) \cdot \mathbb{P}\left(N = \sum_{i=1}^{k} n_i\right) \\
&= \mathbb{P}\left(N_1 = n_1, ..., N_{k-1} = n_{k-1} \middle| N = \sum_{i=1}^{k} n_i, N_k = n_k\right) \cdot \mathbb{P}\left(N_k = n_k \middle| N = \sum_{i=1}^{k} n_i\right) \cdot \mathbb{P}\left(N = \sum_{i=1}^{k} n_i\right) \\
&= \prod_{j=1}^{k-1} \mathbb{P}\left(N_j = n_j \middle| N = \sum_{i=1}^{k} n_i, N_k = n_k, ..., N_{j+1} = n_{j+1}\right) \cdot \mathbb{P}\left(N_k = n_k \middle| N = \sum_{i=1}^{k} n_i\right) \cdot \mathbb{P}\left(N = \sum_{i=1}^{k} n_i\right)
\end{aligned}
$$

$$= \frac{\lambda^{\sum_{i=1}^{k} n_i}}{\left(\sum_{i=1}^{k} n_i\right)!} \cdot e^{-\lambda} \cdot \binom{\sum_{i=1}^{k} n_i}{n_k} p_k^{n_k} (1-p_k)^{\sum_{i=1}^{k-1} n_i} \cdot \prod_{j=1}^{k-1} \binom{\sum_{i=1}^{j} n_i}{n_j} \left(\frac{p_j}{1-\sum_{i=j+1}^{k} p_i}\right)^{n_j} \left(1-\frac{p_j}{1-\sum_{i=j+1}^{k} p_i}\right)^{\sum_{i=1}^{j-1} n_i}$$

$$= \prod_{j=1}^{k} \frac{(\lambda p_j)^{n_j}}{n_j!} e^{-\lambda p_j}$$

$$= \prod_{j=1}^{k} \mathbb{P}(N_j = n_j),$$

where we note that $\mathbb{P}\left(N_1 = n_1 \middle| N = \sum_{i=1}^{k} n_i, N_k = n_k, ..., N_2 = n_2\right) = 1$ in the third equality and use the conditional distribution of multinomials (see Lecture 7 notes) to obtain that

$$N_j \Big| \left(N = \sum_{i=1}^{k} n_i, N_k = n_k, ..., N_{j+1} = n_{j+1}\right) \sim \text{Binomial}\left(\sum_{i=1}^{j} n_i, \frac{p_j}{1-\sum_{i=j+1}^{k} p_i}\right)$$

in the fourth equality. The result follows. □

Remark 2. *The thinning property of Poisson distributions is particular useful when we analyze the count data. For instance, in the analysis of single-cell RNA sequencing data, we have a mapping data matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ for the reads from n cells to p genes, where each entry $X_{ij} \in \mathbb{N}$ of $\boldsymbol{X}$ is the number of reads from ith cell that is mapped to the jth gene [Neufeld et al., 2022]. To split the matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ with count data entries into two independent matrices $\boldsymbol{X}^{train}, \boldsymbol{X}^{test} \in \mathbb{R}^{n\times p}$ under the Poisson assumption on X_{ij} for all $i = 1, ..., n, j = 1, ..., p$, one can leverage the thinning property and implement the following "counting splitting" procedures for all $i = 1, ..., n, j = 1, ..., p$:*

1. *Draw $X_{ij}^{train} | X_{ij} \overset{ind.}{\sim}$ Binomial(X_{ij}, ϵ);*
2. *Take $X_{ij}^{test} = X_{ij} - X_{ij}^{train}$.*

Here, $\epsilon \in (0, 1)$ is some tuning parameter controlling the sample size of $\boldsymbol{X}^{train}$ and its correlation with the original data matrix $\boldsymbol{X}$, which is $\sqrt{\epsilon}$ (recall Problem 2 (c)). By our results in Problem 2, $X_{ij}^{train} \sim$ Poisson$(\epsilon\Gamma_{ij}), X_{ij} \sim$ Poisson$((1-\epsilon)\Gamma_{ij})$ and they are independent when $X_{ij} \sim$ Poisson(Γ_{ij}) for all $i = 1, ..., n, j = 1, ..., p$. This procedure is particularly useful when only one data matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ is available for the tasks of both model estimation and hypothesis testing (or sometimes called post-selection inference [Berk et al., 2013]). While using the same data $\boldsymbol{X}$ for model selection and hypothesis testing (also known as "double dipping" [Kriegeskorte et al., 2009]) fails to control the Type-I error in the inference step, the independence between $\boldsymbol{X}^{train}$ and $\boldsymbol{X}^{test}$ enables researchers to fit the model (or estimate the parameters) on $\boldsymbol{X}^{train}$ and carry out hypothesis testing on $\boldsymbol{X}^{test}$, leading to a valid p-value. More details can be found in Neufeld et al. [2022].

However, Poisson distribution is not the only probability distribution that embraces the thinning property. In the sequel, we will study a similar thinning property for negative binomial (or Pascal) distribution[1]. Recall from Homework 3 (Exercise 2.38 in Casella and Berger 2002) that $Y \sim \text{NegBinomial}(r, p)$ has its probability mass function as:

$$\mathbb{P}(Y = y) = \binom{r+y-1}{y} p^r (1-p)^y$$

for $y = 0, 1, ...$, where $0 < p < 1$ and r is a positive integer. It characterizes the number of observed failures Y before encountering the rth success of a sequence of binary trials with the probability of success as p. Notice that the geometric distribution is a special case of the negative binomial distribution as NegBinomial$(1, p)$.

[1] See https://en.wikipedia.org/wiki/Negative_binomial_distribution

Before introducing the thinning property of the negative binomial distribution, we discuss the connections between negative binomial and Poisson distributions.

Problem 3. *Recall that a Gamma(α, γ) distribution has a density function as $f(x) = \frac{\gamma^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\gamma x}$ with $x > 0$.*

(a) Assume that $Y|\Lambda \sim$ Poisson(Λ) and $\Lambda \sim$ Gamma$\left(r, \frac{r}{\mu}\right)$. Show that the marginal distribution of Y follows a NegBinomial(r, p) distribution with $p = \frac{r}{r+\mu}$.

(b) Show that the negative binomial distribution NegBinomial$\left(r, \frac{r}{r+\lambda}\right)$ converges in distribution to Poisson(λ) when $r \to \infty$.

(c) Calculate the mean and variance of the negative binomial distribution NegBinomial(r, p). Does its variance larger or smaller than its mean?

Proof. (a) We compute the probability mass function of Y by integrating out the prior Λ as:

$$\begin{aligned}
\mathbb{P}(Y = y) &= \int_0^\infty \frac{\lambda^y}{y!} e^{-\lambda} \cdot \frac{(r/\mu)^r}{\Gamma(r)} \lambda^{r-1} e^{-\frac{r\lambda}{\mu}} d\lambda \\
&= \frac{(r/\mu)^r}{y! \cdot \Gamma(r)} \int_0^\infty \lambda^{y+r-1} e^{-\left(1+\frac{r}{\mu}\right)\lambda} d\lambda \\
&= \frac{(r/\mu)^r}{y! \cdot \Gamma(r)} \int_0^\infty \left(\frac{\mu}{\mu+r}\right)^{y+r} v^{y+r-1} e^{-v} dv \quad \text{by } v = \left(\frac{\mu+r}{r}\right)\lambda \\
&= \frac{\Gamma(y+r)}{y! \cdot \Gamma(r)} \left(\frac{r}{\mu}\right)^r \left(\frac{\mu}{\mu+r}\right)^{y+r} \\
&= \binom{y+r-1}{y} \left(\frac{r}{\mu+r}\right)^r \left(\frac{\mu}{\mu+r}\right)^y
\end{aligned}$$

for any $y = 0, 1, ...$

(b) In Homework 3 (Exercise 2.38 in Casella and Berger 2002), we already compute the MGF of $Y \sim$ NegBinomial(r, p) as

$$\begin{aligned}
M_Y(t) &= \mathbb{E}\exp(tY) \\
&= \sum_{y=0}^\infty e^{ty} \binom{r+y-1}{y} p^r (1-p)^y \\
&= p^r \sum_{y=0}^\infty \binom{r+y-1}{y} \left[(1-p)e^t\right]^y \\
&= p^r \left[1 - (1-p)e^t\right]^{-r} \\
&= \left[\frac{p}{1-(1-p)e^t}\right]^r
\end{aligned}$$

when $t < \log\left(\frac{1}{1-p}\right)$, where we use the fact that $\sum_{y=0}^\infty \binom{r+y-1}{y}(1-p)^y = p^{-r}$ to obtain the fourth equality. Thus, when $p = \frac{r}{\lambda+r}$, the MGF of $Y \sim$ NegBinomial$\left(r, \frac{r}{\lambda+r}\right)$ becomes

$$M_Y(t) = \left[\frac{\frac{r}{r+\lambda}}{1 - \left(\frac{r}{r+\lambda}\right) e^t}\right]^r$$

$$
\begin{aligned}
&= \left[\frac{r}{r+\lambda(1-e^t)}\right]^r \\
&= \left[1 - \frac{\lambda(1-e^t)}{r+\lambda(1-e^t)}\right]^r \\
&\to \exp\left[\lambda(e^t - 1)\right]
\end{aligned}
$$

as $r \to \infty$, where the range of t is $\left(-\infty, \log\left(1+\frac{r}{\lambda}\right)\right) \to (-\infty, \infty)$ at the same time. Since $\exp\left[\lambda(e^t-1)\right]$ with $t \in (-\infty,\infty)$ is the MGF of Poisson(λ), we conclude that NegBinomial $\left(r, \frac{r}{\lambda+r}\right)$ converges in distribution to Poisson(λ) as $r \to \infty$.

Another method for arguing this convergence in distribution is to work directly with the probability mass function of NegBinomial $\left(r, \frac{r}{\lambda+r}\right)$ as

$$
\begin{aligned}
\mathbb{P}(Y=y) &= \binom{r+y-1}{y}\left(\frac{r}{r+\lambda}\right)^r\left(\frac{\lambda}{r+\lambda}\right)^y \\
&= \frac{\lambda^y}{y!}\cdot\frac{(r+y-1)\cdots(r+1)r}{(r+\lambda)^y}\cdot\left(\frac{1}{1+\lambda/r}\right)^r \\
&\to \frac{\lambda^y}{y!}\cdot 1\cdot e^{-\lambda},
\end{aligned}
$$

which is the probability mass function of Poisson(λ).

(c) According to the MGF of $Y \sim$ NegBinomial(r,p), we know that

$$
\frac{d}{dt}M_Y(t) = \frac{rp^r(1-p)e^t}{\left[1-(1-p)e^t\right]^{r+1}} \quad \text{and} \quad \frac{d^2}{dt^2}M_Y(t) = \frac{rp^r(1-p)e^t\left[1+r(1-p)e^t\right]}{\left[1-(1-p)e^t\right]^{r+2}}.
$$

By the property of MGF, we obtain that

$$
\mathbb{E}(Y) = \frac{d}{dt}M_Y(t)\Big|_{t=0} = \frac{r(1-p)}{p} \quad \text{and} \quad \mathbb{E}(Y^2) = \frac{d^2}{dt^2}M_Y(t)\Big|_{t=0} = \frac{r(1-p)\left[1+r(1-p)\right]}{p^2}.
$$

Hence, $\mathrm{Var}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}Y)^2 = \frac{r(1-p)}{p^2}$, which is larger than its mean when $p \in (0,1)$. □

Remark 3. *Different from the Poisson distribution, whose mean is always equal to its variance, the negative binomial distribution has its variance larger than its mean. This property of the negative binomial distribution can be utilized to model the overdispersion in the Poisson-based model [Gardner et al., 1995], where the sample variance exceeds the sample mean given the observational data.*

There are some recent studies [Neufeld et al., 2023] about the "count splitting" procedure when each entry of the data matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ does not follow the Poisson distribution but the negative binomial distribution. We will demonstrate through Problem 4 that sampling X_{ij}^{train} via the binomial distribution from the original count data X_{ij} for $i=1,...,n, j=1,...,p$ cannot guarantee the independence between X_{ij}^{train} and $X_{ij}^{test} = X_{ij} - X_{ij}^{train}$. Instead, the thinning property of the negative binomial distribution appears when we sample from X_{ij}^{train} from a beta-binomial distribution; see also Joe [1996].

Problem 4. *Assume that $Z|P \sim$ Binomial(n,P) and $P \sim$ Beta(α,β) for $\alpha,\beta > 0$.*

(a) Show that the marginal distribution of Z follows a BetaBinomial(n,α,β) distribution with its probability mass function as:

$$
\mathbb{P}(Z=z) = \binom{n}{z}\frac{B(z+\alpha, n-z+\beta)}{B(\alpha,\beta)},
$$

where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$ is a Beta function[2].

[2] See https://en.wikipedia.org/wiki/Beta_function.

(b) *Consider $X|Y \sim$ Binomial(Y, ϵ) and $Y \sim$ NegBinomial(r, p) for some $\epsilon \in (0, 1)$. Calculate the (Pearson's) correlation Cor$(X, Y - X)$. Are X and $Y - X$ independent?*

(c) *Suppose again that $Y \sim$ NegBinomial(r, p). If $Z|Y \sim$ BetaBinomial$(Y, \epsilon r, (1-\epsilon)r)$ for some $\epsilon \in (0, 1)$, show that $Z \sim$ NegBinomial$(\epsilon r, p)$, $Y - Z \sim$ NegBinomial$((1-\epsilon)r, p)$ and $Z, Y - Z$ are independent.*

Proof. (a) Recall from Homework 4 (Exercise 4.36 in Casella and Berger 2002) that we have derived the mean and variance of this compound distribution for Z via laws of total expectation and variance. Indeed, the marginal distribution of Z is tractable and we calculate it by integrating out P as:

$$\begin{aligned}\mathbb{P}(Z = z) &= \int_0^1 \binom{n}{z} p^z (1-p)^{n-z} \cdot \frac{1}{B(\alpha, \beta)} p^{\alpha-1}(1-p)^{\beta-1} dp \\ &= \binom{n}{z} \frac{1}{B(\alpha, \beta)} \int_0^1 p^{z+\alpha-1}(1-p)^{n-z+\beta-1} dp \\ &= \binom{n}{z} \frac{B(z+\alpha, n-z+\beta)}{B(\alpha, \beta)}.\end{aligned}$$

(b) By law of total variance, we calculate that

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}\left[\text{Var}(X|Y)\right] + \text{Var}\left[\mathbb{E}(X|Y)\right] \\ &= \mathbb{E}\left[Y\epsilon(1-\epsilon)\right] + \text{Var}(\epsilon Y) \\ &= \epsilon(1-\epsilon)\frac{r(1-p)}{p} + \frac{\epsilon^2 r(1-p)}{p^2},\end{aligned}$$

where we use the fact that $X|Y \sim \text{Binomial}(Y, \epsilon)$ and results in Problem 3 (c). By symmetry, we know that $\text{Var}(Y - X) = \epsilon(1-\epsilon)\frac{r(1-p)}{p} + \frac{(1-\epsilon)^2 r(1-p)}{p^2}$. By the formula in Section 4.2 in Lecture 4 notes, we compute that

$$\begin{aligned}\text{Cov}(X, Y - X) &= \text{Cov}(X, Y) - \text{Var}(X) \\ &= \text{Cov}\left(\mathbb{E}[X|Y], Y\right) - \text{Var}(X) \\ &= \text{Cov}(\epsilon Y, Y) - \text{Var}(X) \\ &= \epsilon \cdot \frac{r(1-p)}{p^2} - \epsilon(1-\epsilon)\frac{r(1-p)}{p} - \frac{\epsilon^2 r(1-p)}{p^2} \\ &= \frac{\epsilon(1-\epsilon)r(1-p)^2}{p^2}.\end{aligned}$$

Finally, the correlation between X and $Y - X$ is

$$\begin{aligned}\text{Cor}(X, Y - X) &= \frac{\text{Cov}(X, Y - X)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} \\ &= \frac{\frac{\epsilon(1-\epsilon)r(1-p)^2}{p^2}}{\sqrt{\frac{r(1-p)\epsilon}{p}\left(1 - \epsilon + \frac{\epsilon}{p}\right) \cdot \frac{r(1-p)(1-\epsilon)}{p}\left(\epsilon + \frac{1-\epsilon}{p}\right)}} \\ &= \frac{(1-p)\sqrt{\epsilon(1-\epsilon)}}{p\sqrt{\left(1 - \epsilon + \frac{\epsilon}{p}\right)\left(\epsilon + \frac{1-\epsilon}{p}\right)}} > 0\end{aligned}$$

when $\epsilon, p \in (0, 1)$. Hence, X and $Y - X$ are not independent.

(c) Given that the MGF of a beta-binomial distribution is of the form of a hypergeometric function and will be intractable to work with, we derive the distribution of Z from its probability mass function as:

$$\begin{aligned}
\mathbb{P}(Z=z) &= \sum_{y=z}^{\infty}\binom{y}{z}\frac{B\left(z+\epsilon r, y-z+(1-\epsilon)r\right)}{B\left(\epsilon r,(1-\epsilon)r\right)}\cdot\binom{r+y-1}{y}p^r(1-p)^y \quad \text{note that we sum over } y \text{ from } z\\
&= \sum_{y=z}^{\infty}\frac{y!}{z!(y-z)!}\cdot\frac{\Gamma(\epsilon r+z)\cdot\Gamma\left(y-z+(1-\epsilon)r\right)\cdot\Gamma(r)}{\Gamma(y+r)\cdot\Gamma(\epsilon r)\cdot\Gamma\left((1-\epsilon)r\right)}\cdot\frac{\Gamma(r+y)}{y!\cdot\Gamma(r)}p^r(1-p)^y\\
&= \frac{\Gamma(\epsilon r+z)}{\Gamma(\epsilon r)\cdot\Gamma\left((1-\epsilon)r\right)\cdot z!}\sum_{y=z}^{\infty}\frac{\Gamma\left(y-z+(1-\epsilon)r\right)}{(y-z)!}p^r(1-p)^y\\
&= \frac{\Gamma(\epsilon r+z)}{\Gamma(\epsilon r)\cdot\Gamma\left((1-\epsilon)r\right)\cdot z!}\sum_{x=0}^{\infty}\frac{\Gamma\left(x+(1-\epsilon)r\right)}{x!}p^r(1-p)^{x+z}\\
&= \frac{\Gamma(\epsilon r+z)}{\Gamma(\epsilon r)\cdot z!}\cdot p^{\epsilon r}(1-p)^z \quad \underbrace{\sum_{x=0}^{\infty}\binom{x+(1-\epsilon)r-1}{x}p^{(1-\epsilon)r}(1-p)^x}_{\text{Summation over the PMF of NegBinomial}((1-\epsilon)r,p)}\\
&= \binom{z+\epsilon r-1}{z}\cdot p^{\epsilon r}(1-p)^z,
\end{aligned}$$

which is the probability mass function (PMF) of NegBinomial$(\epsilon r, p)$ distribution. Notice that we make extensive use of the equality $\Gamma(\alpha+1)=\alpha!$ in our derivations. Likewise, one can follow the above argument to show that the marginal distribution of $Y-Z$ is NegBinomial $((1-\epsilon)r, p)$. Finally, to prove the independence of Z and $Y-Z$, we compute that

$$\begin{aligned}
\mathbb{P}(Z=z, Y-Z=x) &= \mathbb{P}(Z=z, Y=x+z)\\
&= \mathbb{P}(Z=z|Y=x+z)\cdot\mathbb{P}(Y=x+z)\\
&= \binom{x+z}{z}\frac{B\left(z+\epsilon r, x+(1-\epsilon)r\right)}{B\left(\epsilon r,(1-\epsilon)r\right)}\cdot\binom{r+x+z-1}{x+z}p^r(1-p)^{x+z}\\
&= \frac{(x+z)!}{x!z!}\cdot\frac{\Gamma(z+\epsilon r)\cdot\Gamma\left(x+(1-\epsilon)r\right)\cdot\Gamma(r)}{\Gamma(z+x+r)\cdot\Gamma(\epsilon r)\cdot\Gamma\left((1-\epsilon)r\right)}\cdot\frac{\Gamma(r+x+z)}{(x+z)!\cdot\Gamma(r)}\cdot p^r(1-p)^{x+z}\\
&= \frac{\Gamma(z+\epsilon r)\cdot\Gamma\left(x+(1-\epsilon)r\right)}{z!x!\cdot\Gamma(\epsilon r)\cdot\Gamma\left((1-\epsilon)r\right)}\cdot p^{\epsilon r}p^{(1-\epsilon)r}(1-p)^x(1-p)^z\\
&= \binom{z+\epsilon r-1}{z}p^{\epsilon r}(1-p)^z\cdot\binom{x+(1-\epsilon)r-1}{x}p^{(1-\epsilon)r}(1-p)^x\\
&= \mathbb{P}(Z=z)\cdot\mathbb{P}(Y-Z=x).
\end{aligned}$$

It demonstrates that Z and $Y-Z$ are independent. □

Remark 4. *According to our results in Problem 4, if each entry X_{ij} of the count data matrix $\boldsymbol{X}\in\mathbb{R}^{n\times p}$ is assumed to follow a NegBinomial(r,p) distribution, one can still split it into two independent matrices $\boldsymbol{X}^{train},\boldsymbol{X}^{test}\in\mathbb{R}^{n\times p}$ by the following procedures for $i=1,...,n$, $j=1,...,p$ and $\epsilon\in(0,1)$ as:*

1. *Draw $X_{ij}^{train}|X_{ij}\overset{ind.}{\sim}$ BetaBinomial$(X_{ij},\epsilon r,(1-\epsilon)r)$;*
2. *Take $X_{ij}^{test}=X_{ij}-X_{ij}^{train}$.*

This "negative binomial count splitting" procedure is applicable to the overdispersed count data (i.e., *when the sample variance exceeds the sample mean) and facilitates the model estimation and post-selection inference on two independent pieces of the original data.*

References

R. Berk, L. Brown, A. Buja, K. Zhang, and L. Zhao. Valid post-selection inference. *The Annals of Statistics*, pages 802–837, 2013.

G. Casella and R. Berger. *Statistical Inference*. Duxbury advanced series. Thomson Learning, 2nd ed. edition, 2002.

R. Durrett. *Probability: Theory and Examples*, volume 49. Cambridge university press, 5th edition, 2019.

W. Gardner, E. P. Mulvey, and E. C. Shaw. Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological bulletin*, 118(3):392, 1995.

H. Joe. Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability*, 33(3):664–677, 1996.

N. Kriegeskorte, W. K. Simmons, P. S. Bellgowan, and C. I. Baker. Circular analysis in systems neuroscience: the dangers of double dipping. *Nature neuroscience*, 12(5):535–540, 2009.

A. Neufeld, L. L. Gao, J. Popp, A. Battle, and D. Witten. Inference after latent variable estimation for single-cell rna sequencing data. *arXiv preprint arXiv:2207.00554*, 2022.

A. Neufeld, A. Dharamshi, L. L. Gao, and D. Witten. Data thinning for convolution-closed distributions. *arXiv preprint arXiv:2301.07276*, 2023.

M. Perlman. Probability and Mathematical Statistics I (STAT 512 Lecture Notes), 2020. URL https://sites.stat.washington.edu/people/mdperlma/STAT%20512%20MDP%20Notes.pdf.

D. Stoyan, W. S. Kendall, S. N. Chiu, and J. Mecke. *Stochastic geometry and its applications*. John Wiley & Sons, 2013.