Statistics 581, Problem Set 3 Solution
Wellner; 10/21/2009

1. Suppose that X is a random variable with finite fourth moment; $E|X|^4 < \infty$. Then $\mu_4 = E(X-\mu)^4$ is the fourth central moment of X. The ratio $\mu_4/\sigma^4 \equiv \kappa$ is the *kurtosis* of X (or of the distribution function F of X), and $\gamma_2 \equiv \mu_4/\sigma^4 - 3$ is called the *excess of kurtosis*; note that for any $N(\mu, \sigma^2)$ random variable, $\gamma_2 = 0$. Investigate the value of γ_2 for various classical distributions (t_r, uniform, bernoulli, Poission(λ), ...). How big can γ_2 be? How small can γ_2 be?

   Solution: Note that $\mu_4^{1/4} = \{E(X-\mu)^4\}^{1/4} \ge \{E(X-\mu)^2\}^{1/2} = \sigma$ by Liapunov's inequality. Thus $\mu_4/\sigma^4 \ge 1$ always, or $\gamma_2 \equiv \mu_4/\sigma^4 - 3 \ge -2$ with equality if $X = \pm 1$ with probability 1/2 each: then $\mu = 0$, $\sigma^2 = 1$, $\mu_4 = 1$, and $\gamma_2 = -2$.
   For $X \sim N(0,1)$, $\gamma_2 = 0$ since $EX^4 = 3$.
   For $X \sim t_r$, $r > 4$, $\gamma_2 = 6/(r-4) \nearrow \infty$ as $r \searrow 4$; $\gamma_2 \searrow 0$ as $r \nearrow \infty$.
   (Note that since $X \stackrel{d}{=} Z/\sqrt{\chi_r^2/r}$ with Z and χ_r^2 independent, it follows that

   $$\begin{aligned}
   \mu_4 &= E|X|^4 = E|Z|^4 \cdot E(\chi_r^2)^{-2} \cdot r^2 = (3/4)r^2/((r/2-1)(r/2-2)),\\
   \sigma^2 &= E|X|^2 = EZ^2 \cdot E(\chi_r^2)^{-1} = (r/2)/(r/2-1),\\
   \gamma_2 &= \mu_4/\sigma^4 - 3 = 6/(r-4).)
   \end{aligned}$$

   For $X \sim \text{Gamma}(\alpha, \beta)$, $\gamma_2 = 6/\alpha \nearrow \infty$ as $\alpha \searrow 0$.
   For $X \sim \text{Poisson}(\lambda)$, $\gamma_2 = 1/\lambda \nearrow \infty$ as $\lambda \searrow 0$.
   For $X \sim \text{Bernoulli}(p)$, $\gamma_2 = (1-p)^2/p + p^2/(1-p) - 3$ which $= -2$ when $p = 1/2$, and $\nearrow \infty$ when $p \to 0, 1$.
   For $X \sim$ Laplace, or double exponential, with density $2^{-1}\exp(-|x|)$ (as on Ferguson, ACILST page 51), $\mu_4 = 24$ while $\sigma^2 = 2$, so $\gamma_2 = 6$ (not 6.111... as claimed by Ferguson).

2. Suppose that $X_1, \ldots, X_n$ are i.i.d. $N(\theta, \theta^2/r^2)$ where $\theta \in (0, \infty)$ and $r > 0$ is known. Thus the "signal-to-noise ratio" $\mu/\sigma = \theta/\sqrt{\theta^2/r^2} = r$ is a constant.
   (a) Find the score function for θ (for $n = 1$).
   (b) Find the information for θ (for $n = 1$).
   (c) Express the likelihood equation for θ in the form of a polynomial in θ; what is the degree of this polynomial?
   (d) Use standard results to show that the MLE $\widehat{\theta}_n$ of θ is asymptotically normal and find the asymptotic variance.
   (e) Show that $g(x) = \log x$ is a variance stabilizing transformation for the limiting distribution you found in (d).

   Solution: (a) Since

   $$\begin{aligned}
   p_\theta(x) &= \frac{1}{\sqrt{2\pi\theta^2/r^2}} \exp\left(-\frac{1}{2}\frac{(x-\theta)^2}{\theta^2/r^2}\right)\\
   &= \frac{c}{\sqrt{2\pi}\theta} \exp\left(-\frac{r^2}{x}\left(\frac{x}{\theta} - 1\right)^2\right),
   \end{aligned}$$

it follows that

$$\log p_\theta(x) = -\log\theta - \frac{r^2}{2}\left(\frac{x}{\theta}-1\right)^2,$$

and

$$\begin{aligned}\dot\ell_\theta(x) &= -\frac{1}{\theta} - r^2\left(\frac{x}{\theta}-1\right)\left(\frac{-x}{\theta^2}\right)\\ &= \frac{r^2}{\theta^2}\left(\frac{x}{\theta}-1\right)x - \frac{1}{\theta}\\ &= \frac{r^2x^2}{\theta^3} - \frac{r^2 x}{\theta^2} - \frac{1}{\theta}\\ &= \frac{r^2}{\theta}\left(\frac{x}{\theta}-1\right)^2 + \frac{r^2}{\theta}\left(\frac{x}{\theta}-1\right) - \frac{1}{\theta}.\end{aligned}$$

Note that

$$E\dot\ell_\theta(X) = \frac{r^2}{\theta}\cdot\frac{\theta^2}{r^2}\cdot\frac{1}{\theta^2} + 0 - \frac{1}{\theta} = 0.$$

(b) Starting from the 3rd expression for the score function it is easily seen that

$$\ddot\ell_{\theta\theta}(x) = -\frac{3r^2x^2}{\theta^4} + \frac{2r^2x}{\theta^3} + \frac{1}{\theta^2}.$$

and hence

$$\begin{aligned}E(\ddot\ell_{\theta\theta}(X)) &= -\frac{3r^2}{\theta^4}\left(\frac{\theta^2}{r^2}+\theta^2\right) + \frac{2r^2}{\theta^3}\theta + \frac{1}{\theta^2}\\ &= -\frac{r^2}{\theta^2} - \frac{2}{\theta^2} = -(2+r^2)\frac{1}{\theta^2}.\end{aligned}$$

Thus $I(\mu) = (2+r^2)/\theta^2$.
(c) The likelihood equation for θ can be written as

$$\begin{aligned}0 &= \sum_{i=1}^n \dot\ell_\theta(X_i) = \sum_{i=1}^n\left\{\frac{r^2X_i^2}{\theta^3} - \frac{r^2X_i}{\theta^2} - \frac{1}{\theta}\right\}\\ &= \frac{r^2}{\theta^3}\sum_{i=1}^n X_i^2 - \frac{r^2}{\theta^2}\sum_{i=1}^n X_i - \frac{n}{\theta}.\end{aligned}$$

Equivalently, multiplying by θ^3 and dividing by n,

$$0 = r^2 n^{-1}\sum_{i=1}^n X_i^2 - r^2\overline{X}_n\theta - \theta^2.$$

or

$$0 = \theta^2 + r^2\overline{X}_n\theta - r^2\overline{X^2}\ .$$

This is a polynomial equation of degree 2; i.e. a quadratic, and hence has an easy solution:

$$\widehat{\theta} = \frac{-r^2\overline{X} \pm \sqrt{r^4\overline{X}^2 + 4r^2\overline{X^2}}}{2} = \sqrt{r^2\overline{X^2} + r^4\overline{X}^2/4} - r^2\overline{X}/2.$$

by taking the positive root.
(d) Standard MLE theory yields (since this is a sufficiently regular parametric model)

$$\sqrt{n}(\widehat{\theta} - \theta) \to_d N(0, 1/I(\theta)) = N(0, \theta^2/(2 + r^2)).$$

This agrees with a direct calculation: by the multivariate CLT

$$\sqrt{n}\left(\begin{array}{c} \overline{X}_n - \theta \\ \overline{X^2} - (\theta^2 + \theta^2/r^2) \end{array}\right) \to_d N_2(0, \Sigma)$$

where

$$\Sigma = \left(\begin{array}{cc} \theta^2/r^2 & 2\theta^3/r^2 \\ 2\theta^3/r^2 & 2\theta^4(1 + 2r^2)/r^4 \end{array}\right);$$

Here the upper left entry in the matrix Σ is just the variance of $X = X_1$ while the lower right and upper right entries result from

$$\begin{aligned} E\{(X^2 - \sigma^2 - \mu^2)^2\} &= 2\sigma^4 + 4\mu^2\sigma^2 = 2(\theta^2/r^2)^2 + 4\theta^2\theta^2/r^2, \\ E\{(X^2 - \sigma^2 - \mu^2)(X - \mu)\} &= 2\mu\sigma^2 = 2\theta^3/r^2. \end{aligned}$$

Now $\hat{\theta} = g(\overline{X}_n, \overline{X^2})$ where

$$g(u, v) \equiv \sqrt{r^2 v + r^4 u^2/4} - r^2 u/2$$

has derivatives

$$g'(u, v)^T = \left(\begin{array}{c} \frac{\partial}{\partial u} g(u, v) \\ \frac{\partial}{\partial v} g(u, v) \end{array}\right) = \left(\begin{array}{c} 2^{-1}(r^2 v + r^4 u^2/4)^{-1/2}((r^4 u/2) - r^2/2 \\ 2^{-1}(r^2 v + r^4 u^2/4)^{-1/2}(r^2) \end{array}\right).$$

Evaluating these at $(\theta, \theta^2(r^2 + 1)/r^2$ yields

$$g'(\theta, \theta^2(r^2 + 1)/r^2)^T = \frac{r^2/2}{1 + r^2/2}\left(\begin{array}{c} -1 \\ \theta^{-1} \end{array}\right).$$

Hence, by the g'-theorem

$$\sqrt{n}(\hat{\theta}_n - \theta) \to_d g'Y \sim N(0, g'\Sigma(g')^T) = N(0, \theta^2/(2 + r^2))$$

after a bit of algebra. Another way to proceed would be to rewrite $\hat{\theta}_n$ as

$$\begin{aligned} \hat{\theta}_n &= \sqrt{r^2\overline{X^2} + r^4\overline{X}^2/4} - r^2\overline{X}/2 \\ &= \sqrt{r^2(\overline{X^2} - \overline{X}^2) + (r^4/4 + r^2)\overline{X}^2} - r^2\overline{X}/2 \\ &\equiv \tilde{g}(\overline{X}, S^2), \end{aligned}$$

and then use the joint asymptotic distribution of problem 3(a) below combined with the delta-method.
(e) Since $g'(\theta)^2\theta^2 = 1$ implies $g'(\theta) = 1/\theta$ and hence $g(\theta) = \log\theta$, it is easily seen that $g(x) = \log x$ is a variance stabilizing transformation. We conclude that

$$\sqrt{n}(\log\widehat{\theta}_n - \log\theta) \to_d N(0, 1/(2+r^2)).$$

3. Suppose that $X_1, X_2, \ldots$ are i.i.d. (μ, σ^2) with $\mu_4 < \infty$. Let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ and $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ be the sample mean and sample variance respectively.
(a) Show that

$$\sqrt{n}\left(\begin{array}{c} \overline{X}_n - \mu \\ S_n^2 - \sigma^2 \end{array}\right) \to_d \underline{Z} \sim N_2(0, \Sigma)$$

where

$$\left(\begin{array}{cc} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{array}\right).$$

(b) Suppose $\mu > 0$. Use (a) to find the limiting distribution of the sample *signal to noise ratio* $R_n \equiv \overline{X}_n/S_n$; i.e. show that $\sqrt{n}(R_n - r) \to_d N(0, V^2)$ with $r \equiv \mu/\sigma$ and find V^2.

Solution: (a) Since $S_n^2 = n^{-1}\sum_{i=1}^n (X_i - \mu)^2 + o_p(1/\sqrt{n})$, we have

$$\begin{aligned} \sqrt{n}\left(\begin{array}{c} \overline{X}_n - \mu \\ S_n^2 - \sigma^2 \end{array}\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\begin{array}{c} X_i - \mu \\ (X_i - \mu)^2 - \sigma^2 \end{array}\right) + o_p(1) \\ &\to_d \underline{Z} \sim N_2(0, \Sigma) \end{aligned}$$

by the multivariate CLT where Σ is as given above.
(b) The function $g(u, v) = u/\sqrt{v}$ is differentiable at points (u, v) with $v \neq 0$, and the derivative is $\nabla g(u, v) = (1/\sqrt{v}, u(-1/2)v^{-3/2})$ so that $\nabla g(\mu, \sigma^2) = (1/\sigma, (-1/2)\mu\sigma^{-3}) = (1/\sigma)(1, -(1/2)\mu/\sigma^2))$. Hence it follows from the delta method (g' theorem) that

$$\begin{aligned} \sqrt{n}(R_n - r) &= \sqrt{n}\left(\frac{\overline{X}_n}{S_n} - \frac{\mu}{\sigma}\right) \\ &= \sqrt{n}(g(\overline{X}_n, S_n^2) - g(\mu, \sigma^2)) \\ &\to_d \nabla g \cdot \underline{Z} \sim N(0, \nabla g^T \Sigma \nabla g) \end{aligned}$$

and it is easy to calculate that

$$\begin{aligned} \nabla g^T \Sigma \nabla g &= \frac{1}{\sigma^4}\left\{\sigma^4 - \mu\mu_3 + \frac{1}{4}d^2(\mu_4 - \sigma^4)\right\} \\ &= 1 - r\gamma_1 + \frac{1}{4}r^2(2 + \gamma_2) \end{aligned}$$

where $\gamma_1 \equiv \mu_3/\sigma^3$ and $\gamma_2 \equiv \mu_4/\sigma^4 - 3$. Note that when the X_i's are normal (so $\gamma_1 = \gamma_2 = 0$), this reduces to $1 + r^2/2$. Thus under normality we have

$$\sqrt{n}(g(R_n) - g(r)) \to_d N(0,1)$$

if $g(x) \equiv \sqrt{2}\text{arcsinh}(x/\sqrt{2})$.

4. Ferguson, ACILST, page 34, problem 1(b), modified slightly.
Suppose that $X_1, \ldots, X_n$ is a sample from the Poisson distribution with parameter $\lambda > 0$: $p_k(\lambda) = P(X_1 = k) = \exp(-\lambda)\lambda^k/k!$, $k = 0, 1, \ldots$. Let J be a fixed positive integer (e.g. $J = 5$), and consider the following two estimators of $\underline{p} = (p_0, p_1, \ldots, p_J) \in \mathbb{R}^{J+1}$:
$$\widehat{\underline{p}} = (p_0(\hat\lambda), p_1(\hat\lambda), \ldots, p_J(\hat\lambda))$$
where $\hat\lambda = \overline{X}_n = n^{-1}\sum_1^n X_i$, and
$$\widetilde{\underline{p}} = n^{-1}\sum_{i=1}^n (1_{[X_i=0]}, 1_{[X_i=1]}, \ldots, 1_{[X_i=J]}).$$
(a) What is the joint asymptotic distribution of
$$\sqrt{n}\begin{pmatrix} \widehat{\underline{p}} - \underline{p} \\ \widetilde{\underline{p}}_n - \underline{p} \end{pmatrix}$$
as a vector in $\mathbb{R}^{2(J+1)}$? Is the resulting joint distribution nondegenerate? (Why or why not?)
(b) Consider $p_j(\lambda) \equiv P_\lambda(X_1 = j)$. What is the joint asymptotic distribution of $\hat p_j \equiv p_j(\hat\lambda_n)$ and $\tilde p_j$ where $\hat\lambda_n = \overline{X}_n$ and $\tilde p_j = n^{-1}\sum_{i=1}^n 1_{[X_i=j]}$?
(c) Compute the ratio of the asymptotic variances of the two estimators $\hat p_j$ and $\tilde p_j$ of p_j. How does this ratio behave as a function of j?
(d) Which estimator would you prefer if the Poisson model (assumption) holds? Which estimator would you prefer if the Poisson model (assumption) fails?

Solution: (a) First consider the vectors $\underline{Y}_n \equiv \sqrt{n}(\overline{X}_n - \lambda, (\widetilde{p}_n - p)^T)^T$ in $\mathbb{R}^{J+2}$. Note that

$$\underline{Y}_n = \sqrt{n}\underline{W}_n$$

where the random vectors

$$\underline{W}_i \equiv \begin{pmatrix} X_i - \lambda \\ 1_{\{0\}}(X_i) - p_0 \\ \vdots \\ 1_{\{J\}}(X_i) - p_J \end{pmatrix}$$

are i.i.d. (since the X_i's are i.i.d.) with $E\underline{W}_1 = \underline{0}$ and covariance matrix

$$\Sigma = E(\underline{W}_1\underline{W}_1^T) = \begin{pmatrix} \lambda & (0-\lambda)p_0 & \cdots & (J-\lambda)p_J \\ (0-\lambda)p_0 & p_0(1-p_0) & \cdots & -p_0p_J \\ \vdots & & \cdots & \vdots \\ (J-\lambda)p_J & -p_0p_J & \cdots & p_J(1-p_J) \end{pmatrix} \equiv \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Thus by the multivariate CLT,

$$\underline{Y}_n = \sqrt{n}\underline{W}_n \to_d \underline{Y} \sim N_{J+2}(0, \Sigma).$$

Now the vector of interest,

$$\sqrt{n}((\widehat{\underline{p}} - \underline{p})^T, (\widetilde{\underline{p}}_n - \underline{p})^T)^T$$

can be written as

$$\sqrt{n}\begin{pmatrix} \widehat{\underline{p}} - \underline{p} \\ \widetilde{\underline{p}}_n - \underline{p} \end{pmatrix} = \sqrt{n}(g(\hat{\lambda}, \widetilde{\underline{p}}_n) - g(\lambda, \underline{p}))$$

where $g : \mathbb{R}^+ \times \mathbb{R}^{J+1} \subset \mathbb{R}^{J+2} \to \mathbb{R}^{2(J+1)}$ is defined by $g(u, \underline{v}) = (p_0(u), \ldots, p_J(u), v_1, \ldots, v_{J+1})$ for $u \in \mathbb{R}^+$, $v \in \mathbb{R}^{J+1}$. To apply the delta method we calculate the $2(J+1) \times (J+2)$-derivative matrix g':

$$g'(u,v) = \begin{pmatrix} \frac{\partial}{\partial u}p_0(u) & 0 & \cdots & \cdots & 0 \\ \vdots & 0 & \vdots & \cdots & 0 \\ \frac{\partial}{\partial u}p_J(u) & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \equiv \begin{pmatrix} g'_{11} & 0 \\ 0 & I \end{pmatrix} \equiv \begin{pmatrix} g'_{11} & g'_{12} \\ g'_{21} & g'_{22} \end{pmatrix},$$

so that

$$g' \equiv g'(\lambda, p) = \begin{pmatrix} p'_0(\lambda) & 0 & \cdots & \cdots & 0 \\ \vdots & 0 & \vdots & \cdots & 0 \\ p'_J(\lambda) & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \equiv \begin{pmatrix} g'_{11} & 0 \\ 0 & I \end{pmatrix}$$

where $g'_{11,j} = p_{j-1}(\lambda) - p_j(\lambda) = p_j(\lambda)(j/\lambda - 1)$. Thus it follows from the g' theorem that

$$\begin{aligned} \sqrt{n}\begin{pmatrix} \widehat{\underline{p}} - \underline{p} \\ \widetilde{\underline{p}}_n - \underline{p} \end{pmatrix} &= \sqrt{n}(g(\hat{\lambda}, \widetilde{\underline{p}}_n) - g(\lambda, \underline{p})) \\ &\to_d g'\underline{Y} \sim N_{2J+2}(0, g'\Sigma g) \end{aligned}$$

with

$$g'\Sigma(g')^T = \begin{pmatrix} g'_{11}(g'_{11})^T\lambda & g'_{11}\Sigma_{12} \\ \Sigma_{21}(g'_{11})^T & \mathrm{diag}(\underline{p}_J) - \underline{p}_J\underline{p}_J^T \end{pmatrix}.$$

where $\underline{p}_J \equiv (p_0, \ldots, p_J)^T$. This matrix is singular: note that if $d = (d_0, \ldots, d_J, 0, \ldots, 0) \in \mathbb{R}^{2J+2}$ is chosen so that

$$d^T g'_{11} = \sum_{j=0}^{J} d_j (p_{j-1} - p_j) = \sum_{j=0}^{J} d_j p_j(\lambda)(\frac{j}{\lambda} - 1) = 0,$$

(which is easily possible by taking, e.g. $d_0 = d_j = (p_{J-1} - p_0)/(p_{J-1} - p_0 - p_J)$ and $d_1 = d_2 = \cdots = d_{J-1} = 1$, assuming that $p_{J-1} - p_0 - p_J \neq 0$), we find that $d^T \Sigma d = 0$. Thus the random vector $g'\underline{Y}$ has a degenerate normal distribution the rank of Σ is no more than $J + 2$. (Note that $g'_{11,j} = p_{j-1}(\lambda) - p_j(\lambda) = 0$ if $j = \lambda$, which could reduce the rank of Σ to $J + 1$.)
(b) From (a) it follows that the joint asymptotic distribution of $\hat{p}_j \equiv p_j(\hat{\lambda}_n)$ and $\tilde{p}_j$ is given by

$$\sqrt{n} \begin{pmatrix} \hat{p}_j - p_j \\ \tilde{p}_j - p_j \end{pmatrix} \rightarrow_d (Y_j, Y_{j+(J+1)})^T \sim N_2(0, \Sigma_{j,j+J+1})$$

where

$$\Sigma_{j,j+J+1} = \begin{pmatrix} \lambda(1 - j/\lambda)^2 p_j^2(\lambda) & c_j \\ c_j & p_j(1 - p_j) \end{pmatrix}.$$

where $c_j = (p_{j-1} - p_j)(j - \lambda)p_j = (j - \lambda)^2 p_j^2/\lambda$. Thus the ratio of asymptotic variances is

$$\begin{aligned} ARE_{\hat{p}_j,\tilde{p}_j}(\lambda, j) &= \frac{\lambda(1 - j/\lambda)^2 p_j^2(\lambda)}{p_j(\lambda)(1 - p_j(\lambda))} \\ &= \frac{(j - \lambda)^2 p_j(\lambda)}{\lambda(1 - p_j(\lambda))}. \end{aligned}$$

As a function of j for a fixed λ this typically decreases to j's near $\lfloor \lambda \rfloor$, then increases before decreasing again as the $\lambda^{j-1}/j!$ term from combining the numerator and denominator wins out against the quadratic $(j - \lambda)^2$.
(c) When the Poisson model is true, we would clearly prefer the model based estimators $\hat{p}_j$ since they have smaller variance. When the Poisson model fails, the advantage of the raw proportion estimators is that they remain consistent when the Poisson model fails.

5. Let $X_{n1}, \ldots, X_{nn}$ be independent, $X_{nk} \sim \text{Bernoulli}(p_{nk})$, and let $Y_n \sim \text{Poisson}(\sum_{k=1}^n p_{nk})$. Let P_n be the distribution of $\sum_{k=1}^n X_{nk}$ and let Q_n be the distribution of Y_n. Show that

$$d_{TV}(P_n, Q_n) \equiv \sup_{A \in \mathcal{B}} |P(S_n \in A) - P(Y_n \in A)| \leq \sum_{k=1}^n p_{nk}^2.$$

Note that when $p_{nk} = p_n \to 0$ for all k and $np_n \to \lambda$, then $\sum_{k=1}^n p_{nk}^2 = np_n^2 = (np_n)^2/n = O(n^{-1})$.

[Hint: construct S_n and Y_n on a common probability space as follows: let $T_{nk} \sim$ Poisson(p_{nk}), $k = 1, \ldots, n$ be independent, and let $Z_{nk} \sim$ Bernoulli$(1-(1-p_{nk})e^{p_{nk}})$, $k = 1, \ldots, n$ be independent and independent of the T_{nk}'s. Define

$$X_{nk} = 1_{[T_{nk} \geq 1]} + 1_{[T_{nk}=0]} 1_{[Z_{nk}=1]}.$$

Set $S_n = \sum_{k=1}^n X_{nk}$, $Y_n = \sum_{k=1}^n T_{nk}$. Check that $X_{nk} \sim$ Bernoulli(p_{nk}), $Y_n \sim$ Poisson$(\sum_1^n p_{nk})$, and

$$\begin{aligned} P(T_{nk} = 0, X_{nk} = 1) &= e^{-p_{nk}} - (1 - p_{nk}) \\ P(T_{nk} \geq 1, X_{nk} = 0) &= 0 \\ P(T_{nk} \geq 2) &= 1 - e^{-p_{nk}} - p_{nk} e^{-p_{nk}}. \end{aligned}$$

Show that

$$d_{TV}(P_n, Q_n) \leq P(S_n \neq Y_n) \leq \sum_{k=1}^n P(X_{nk} \neq T_{nk}) \leq \sum_{k=1}^n p_{nk}^2.]$$

Solution: We first verify the computations: with $T_{nk} \sim$ Poisson(p_{nk}) and $Z_{nk} \sim$ Bernoulli$(1 - (1 - p_{nk})e^{p_{nk}})$ all independent,

$$\begin{aligned} &P(T_{nk} = 0, X_{nk} = 1) = P(T_{nk} = 0, Z_{nk} = 1) = P(T_{nk} = 0)P(Z_{nk} = 1) \\ &\quad = e^{-p_{nk}}(1 - (1 - p_{nk})e^{p_{nk}}) = e^{-p_{nk}} - (1 - p_{nk}), \\ &P(T_{nk} \geq 1, X_{nk} = 0) = P(X_{nk} = 1, X_{nk} = 0) = 0, \\ &P(T_{nk} \geq 2) = 1 - P(T_{nk} \leq 1) = 1 - e^{-p_{nk}} - p_{nk} e^{-p_{nk}}, \\ &P(X_{nk} = 1) = P(T_{nk} \geq 1) + P(T_{nk} = 0, Z_{nk} = 1) \\ &\quad = 1 - e^{-p_{nk}} + e^{-p_{nk}} - (1 - p_{nk}) = p_{nk}. \end{aligned}$$

Thus $X_{nk} \sim$ Bernoulli(p_{nk}), $S_n \equiv \sum_{k=1}^n X_{nk}$, and $T_n \equiv \sum_{k=1}^n T_{nk} \sim$ Poisson$(\sum_1^n p_{nk})$.

Now let $A \subset \mathbb{N} = \{0, 1, 2, \ldots\}$. Then

$$\begin{aligned} P_n(A) - Q_n(A) &= P(S_n \in A, S_n = T_n) + P(S_n \in A, S_n \neq T_n) \\ &\quad - P(T_n \in A, S_n = T_n) - P(T_n \in A, S_n \neq T_n) \\ &= P(S_n \in A, S_n = T_n) - P(T_n \in A, S_n = T_n) \\ &\quad + P(S_n \in A, S_n \neq T_n) - P(T_n \in A, S_n \neq T_n) \\ &\leq P(S_n \in A, S_n = T_n) - P(S_n \in A, S_n = T_n) + P(S_n \neq T_n) \\ &= P(S_n \neq T_n). \end{aligned}$$

Similarly,

$$Q_n(A) - P_n(A) \leq 0 + P(S_n \neq T_n),$$

and hence

$$|P_n(A) - Q_n(A)| \leq P_n(S_n \neq T_n)$$

for all subsets A of $\mathbb{N}$. Therefore it follows that

$$\begin{aligned}
d_{TV}(P_n, Q_n) &\le P(S_n \ne T_n) \le \sum_{k=1}^n P(X_{nk} \ne T_{nk}) \\
&= \sum_{k=1}^n \{P(T_{nk} = 0,\ X_{nk} = 1) + P(T_{nk} \ge 1, X_{nk} = 0) + P(T_{nk} \ge 2)\} \\
&= \sum_{k=1}^n \{e^{-p_{nk}} - (1 - p_{nk}) + 0 + 1 - e^{-p_{nk}} - p_{nk} e^{-p_{nk}}\} \\
&= \sum_{k=1}^n p_{nk}(1 - e^{-p_{nk}}) \le \sum_{k=1}^n p_{nk}^2
\end{aligned}$$

where the last inequality follow from $1 - e^{-x} \le x$ for all $x \ge 0$. [Notes: this bound via a coupling argument is from Hodges and LeCam (1960), Ann. Math. Statist. 31. For the Stein-Chen inequality giving a still tighter bound for large values of $\lambda_n = \sum_1^n p_{nk}$, see Barbour, Holst, and Janson (1992), *Poisson Approximation*.