

# Homework 1

Juliette Franqueville

September 19, 2022

## (1) Prove Slutsky's theorem.

Slutsky's theorem states that

- $x_n \xrightarrow{P} x \Rightarrow x_n \xrightarrow{D} x$
- $x_n \xrightarrow{D} x_0 \Rightarrow x_n \xrightarrow{P} x_0$
- if  $x_n \xrightarrow{D} x$  and  $z_n \xrightarrow{P} z_0$ , then

$$x_n + z_n \xrightarrow{D} x + z_0$$

$$x_n z_n \xrightarrow{D} x z_0$$

$$x_n / z_n \xrightarrow{D} x / z_0$$

We will prove the three lemmas in order.

First, we have to show that if  $x_n \xrightarrow{P} x$ ,  $\lim_{n \rightarrow \infty} Pr(x_n \leq a) = Pr(x \leq a)$  for any  $a$ .

We show, using basic probability rules, that for  $\epsilon > 0$

$$\begin{aligned} Pr(x_n \leq a) &= Pr(x_n \leq a, x \leq a + \epsilon) + Pr(x_n \leq a, x > a + \epsilon) \\ &\leq Pr(x \leq a + \epsilon) + Pr(x_n - x > a - x, a - x < -\epsilon) \\ &\leq Pr(x \leq a + \epsilon) + Pr(x_n - x < -\epsilon) \\ &\leq Pr(x \leq a + \epsilon) + Pr(x_n - x < -\epsilon) + Pr(x_n - x > \epsilon) \\ &\leq Pr(x \leq a + \epsilon) + Pr(|x_n - x| > \epsilon) \end{aligned}$$

So,

$$\begin{aligned} Pr(x_n \leq a) &\leq Pr(x \leq a + \epsilon) + Pr(|x_n - x| > \epsilon) \\ Pr(x_n \leq a - \epsilon) &\leq Pr(x \leq a) + Pr(|x_n - x| > \epsilon) \\ Pr(x_n \leq a - \epsilon) - Pr(|x_n - x| > \epsilon) &\leq Pr(x_n \leq a) \leq Pr(x_n \leq a + \epsilon) + Pr(|x_n - x| > \epsilon) \end{aligned}$$

Taking the limit as  $n \rightarrow \infty$ , we find that  $F_x(a - \epsilon) \leq \lim_{n \rightarrow \infty} Pr(x_n \leq a) \leq F_x(a + \epsilon)$ , so (from the squeeze theorem),  $\lim_{n \rightarrow \infty} Pr(x_n \leq a) = Pr(x \leq a)$  as  $\epsilon \rightarrow 0$ .

Secondly, we have to show that if  $x_n \xrightarrow{D} x_0$ ,  $\lim_{n \rightarrow \infty} P(|x_n - x| > \epsilon) = 0$ . We have:

$$\begin{aligned} \lim_{n \rightarrow \infty} P(|x_n - x_0| \leq \epsilon) &= \lim_{n \rightarrow \infty} P(-\epsilon \leq x_n - x_0 \leq \epsilon) \\ &= \lim_{n \rightarrow \infty} P(x_n - x_0 \leq \epsilon) - \lim_{n \rightarrow \infty} P(x_n - x_0 \leq -\epsilon) \\ &= \lim_{n \rightarrow \infty} P(x_n \leq \epsilon + x_0) - \lim_{n \rightarrow \infty} P(x_n \leq x_0 - \epsilon) \\ &= \lim_{n \rightarrow \infty} F_{x_n}(\epsilon + x_0) - \lim_{n \rightarrow \infty} F_{x_n}(x_0 - \epsilon) \\ &= 1 - 0 = 1 \end{aligned}$$

Or, equivalently,  $\lim_{n \rightarrow \infty} P(|x_n - x_0| > \epsilon) = 0$ ; which is the definition of convergence in probability. To go from the second to last to the last line, we used the fact that  $x_n \xrightarrow{D} x_0$ .

We know that if  $z_n \xrightarrow{P} z_0$ ,  $z_n \xrightarrow{D} z_0$ . Lastly, we show that  $(x_n, z_n) \xrightarrow{D} (x, z_0)$ . If  $(x_n, z_0) \xrightarrow{D} (x, z_0)$  and  $(x_n, z_n) - (x_n, z_0) \xrightarrow{P} 0$ , then  $(x_n, z_n) \xrightarrow{D} (x, z_0)$ .  $(x_n, z_0) \xrightarrow{D} (x, z_0)$  follows from the continuous mapping theorem, and  $(x_n, z_n) - (x_n, z_0) = z_n - z_0 \xrightarrow{P} 0$ , so  $(x_n, z_n) \xrightarrow{D} (x, z_0)$ . We then simply use the continuous mapping theorem:

$$\begin{aligned} x_n &\xrightarrow{D} x \\ g(x_n) &\xrightarrow{D} g(x) \end{aligned}$$

with  $g(x, y) = x + y$ ,  $g(x, y) = xy$ , and  $g(x, y) = x/y$  to prove the last part of Slutsky's theorem.

**(2) For  $y \sim \text{Poisson}(\lambda)$ , show that  $E(y) = \text{var}(y) = \lambda$ . Method of moments suggests  $\bar{y}_n$ , the sample mean, as well as  $s_n^2$ , the sample variance, could both be reasonable estimators of  $\lambda$ . Which one would you prefer? Why?**

It is shown in question (7) that the MGF of the Poisson distribution is:

$$MGF_y(t) = e^{\lambda(e^t - 1)}$$

We use the first derivative of the MGF,  $\lambda e^t e^{\lambda(e^t - 1)}$ , to find that  $E(y) = \lambda$ . To find the variance, we use  $\text{var}(y) = E(y^2) - E(y)^2$ .  $E(y^2)$  is the second moment, which we calculate using the second derivative of the MGF,  $\lambda e^t e^{\lambda(e^t - 1)} + \lambda^2 e^{2t} e^{\lambda(e^t - 1)}$ , evaluated at 0 which is  $\lambda + \lambda^2$ . So,  $\text{var}(y) = E(y^2) - E(y)^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$ .

It seems more natural to use the sample mean as an estimator of  $\lambda$  than the sample variance, although both may be used. Also, since the MLE for Poisson is the sample mean, the sample mean may be preferable as it will maximize the likelihood of the data.

**(3) For  $(x_1, y_1), \dots, (x_n, y_n) \sim f_{x,y}$  with finite second order moments. Show that the sample correlation coefficient  $r_n$  converges in probability to the population correlation coefficient  $\rho$ .**

$\rho$  can be estimated using  $r_n$ :

$$\begin{aligned} r_n &= \frac{(n-1)^{-1} \sum (x_j - \bar{x})(y_j - \bar{y})}{[(n-1)^{-1} \sum (x_j - \bar{x})^2 (n-1)^{-1} \sum (y_j - \bar{y})^2]^{1/2}} \\ &= \frac{(n-1)^{-1} [\sum x_j y_j - \bar{y} \sum x_j - \bar{x} \sum y_j + \bar{x} \bar{y}]}{[(n-1)^{-1} \sum (x_j - \bar{x})^2 (n-1)^{-1} \sum (y_j - \bar{y})^2]^{1/2}} \end{aligned}$$

The sample variances  $(n-1)^{-1} \sum (x_j - \bar{x})^2$  and  $(n-1)^{-1} \sum (y_j - \bar{y})^2$  converge to the population variances  $\text{var}(x)$  and  $\text{var}(y)$ , respectively (this was shown in class). Using continuous mapping theorem,  $[(n-1)^{-1} \sum (x_j - \bar{x})^2]^{1/2}$  and  $[(n-1)^{-1} \sum (y_j - \bar{y})^2]^{1/2}$  converge to  $\text{var}(x)^{1/2}$  and  $\text{var}(y)^{1/2}$ , respectively.

In the numerator, we know that  $n^{-1} \sum x_j$  converges in probability to  $E(x)$  (from WLLN), so  $(n-1)^{-1} \sum x_j = [(n-1)^{-1}/n^{-1}] n^{-1} \sum x_j$  also converges to  $E(x)$ . Likewise,  $(n-1)^{-1} \sum y_j$  converges to  $E(y)$  and  $(n-1)^{-1} \sum x_j y_j$  converges to  $E(xy)$ .  $\bar{x}$  and  $\bar{y}$  converge to  $E(x)$  and  $E(y)$ , respectively. Putting this all together and using the fact that if  $X_n \xrightarrow{P} X$  and  $Y_n \xrightarrow{P} Y$ , then  $X_n + Y_n \xrightarrow{P} X + Y$ ,  $X_n Y_n \xrightarrow{P} XY$  and  $X_n/Y_n \xrightarrow{P} X/Y$  (this follows from  $X_n \xrightarrow{P} X$ ,  $Y_n \xrightarrow{P} Y \Rightarrow (X_n, Y_n) \xrightarrow{P} (X, Y)$  and applying the continuous mapping theorem), we get that

$$\begin{aligned} r_n &\xrightarrow{P} \frac{E(xy) - E(y)E(x) - E(x)E(y) + E(x)E(y)}{[\text{var}(x)\text{var}(y)]^{1/2}} \\ r_n &\xrightarrow{P} \frac{E(xy) - E(y)E(x)}{[\text{var}(x)\text{var}(y)]^{1/2}} \\ r_n &\xrightarrow{P} \rho \end{aligned}$$

**(4) Notations having their usual significance, for  $y_1, \dots, y_n \text{ iid} \sim \text{Ga}(\alpha, \beta)$ , a method of moment estimator of  $\alpha$  is  $\hat{\alpha}_n = \bar{y}_n^2/s_n^2$ . Using the multivariate delta method, show that  $SE(\hat{\alpha}_n) \approx \sqrt{2\alpha(\alpha+1)/n}$  for large values of  $n$ .**

The following equations were proven in class:

$$\begin{aligned} \text{var}(\bar{y}_n) &= \frac{\alpha}{\beta^2 n} \\ \text{var}(s_n^2) &= \frac{6\alpha + 2\alpha^2}{\beta^4 n} \\ \text{cov}(\bar{y}_n, s_n^2) &= \frac{2\alpha}{\beta^3 n} \end{aligned}$$

We have

$$\begin{aligned}
\text{var}(t_{1,n}) &= \text{var}(\bar{y}_n) \\
&= \frac{\text{var}(y_n)}{n} \\
&= \frac{\alpha}{n\beta^2}
\end{aligned}$$

and,

$$\begin{aligned}
\text{var}(t_{2,n}) &= \text{var}(s_n^2) \\
&= \frac{6\alpha + 2\alpha^2}{\beta^4 n}
\end{aligned}$$

We can then form the covariance matrix:

$$\Sigma = \begin{pmatrix} \frac{\alpha}{\beta^2} & \frac{2\alpha}{\beta^3} \\ \frac{2\alpha}{\beta^3} & \frac{6\alpha + 2\alpha^2}{\beta^4} \end{pmatrix}$$

Using the delta method, we have:

$$\sqrt{n}[h(t_{1,n}, t_{2,n}) - h(\theta_1 - \theta_2)] \xrightarrow{D} \mathcal{N}(0, h'(\theta_1 - \theta_2)^T \Sigma h'(\theta_1 - \theta_2))$$

We have

$$\begin{aligned}
E(t_{1,n}) &= E(\bar{y}_n) \\
&= \frac{\alpha}{\beta}
\end{aligned}$$

$$\begin{aligned}
E(t_{1,2}) &= E(s_n^2) \\
&= \frac{\alpha}{\beta^2}
\end{aligned}$$

Then,

$$\begin{aligned}
\frac{\partial h}{\partial \theta_1} &= \frac{2\theta_1}{\theta_2} \\
&= \frac{2\alpha}{\beta} \frac{\beta^2}{\alpha} \\
&= 2\beta
\end{aligned}$$

$$\begin{aligned}
\frac{\partial h}{\partial \theta_2} &= \frac{-\theta_1^2}{\theta_2} \\
&= -\frac{\alpha^2}{\beta^2} \frac{\beta^4}{\alpha^2} \\
&= -\beta^2
\end{aligned}$$

We then find

$$\begin{aligned} \text{var}(\sqrt{n}\hat{\alpha}_n) &= h'(\theta_1 - \theta_2)^T \Sigma h'(\theta_1 - \theta_2) \\ &= 2\alpha(\alpha + 1) \\ \text{var}(\hat{\alpha}_n) &= \frac{2\alpha(\alpha + 1)}{n} \end{aligned}$$

We then obtain the standard error:

$$\begin{aligned} SE(\hat{\alpha}_n) &= [\text{var}(\hat{\alpha}_n)]^{1/2} \\ &= \sqrt{\frac{2\alpha(\alpha + 1)}{n}} \end{aligned}$$

(5) Now fix the values of  $\alpha$  and  $\beta$ . For your chosen values of  $\alpha$  and  $\beta$ , draw a random sample of size  $n = 50$  from a  $\text{Ga}(\alpha, \beta)$  distribution. Using the method of moments and assuming  $\alpha$  and  $\beta$  to now be unknown, estimate  $\alpha$  and  $\beta$ . Plot the histogram of the samples, superimposed with the true density and the estimated density. Repeat the above procedure  $B = 50, 500$  and  $1000$  times. Plot a histogram of  $\sqrt{n} \frac{(\hat{\alpha}_n - \alpha)}{\sqrt{2\hat{\alpha}_n(\hat{\alpha}_n + 1)}}$  for each of the above values of  $B$ . In each case, superimpose a  $\text{Normal}(0,1)$  distribution over the histogram.

The figures below shows the estimated and actual densities and  $\sqrt{n} \frac{(\hat{\alpha}_n - \alpha)}{\sqrt{2\hat{\alpha}_n(\hat{\alpha}_n + 1)}}$  and the  $N(0,1)$  density as  $B$  increases.

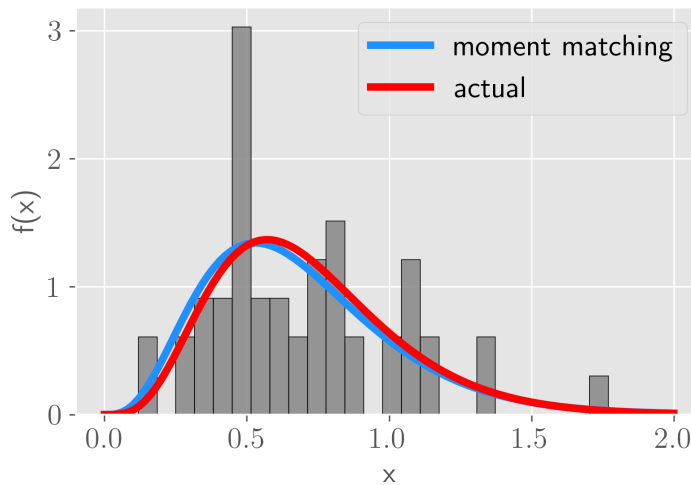


Figure 1: Estimated and actual densities with  $\alpha = 5, \beta = 7$

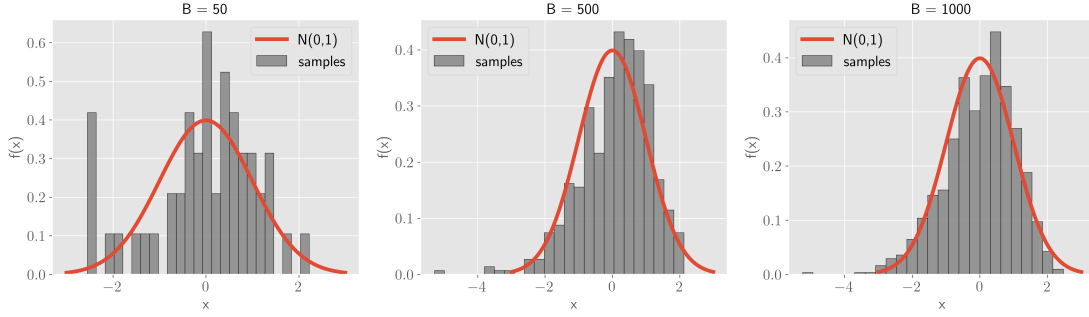


Figure 2: Samples and  $N(0,1)$

The first figure shows that the density estimated using moment matching is close to the actual one. Increasing the number of samples would improve the estimated density. The second figure confirms that the standard error derived for the estimator  $\hat{\alpha}_n$  is  $\sqrt{\frac{2\alpha(\alpha+1)}{n}}$ ; as  $B$  increases the samples look more and more like the  $N(0,1)$  density.  $n = 50$  seems to be sufficiently high.

**(6) For a  $\text{Normal}(\mu, \sigma^2)$  distribution, show that the MGF is  $M(t) = \exp(\mu t + \sigma^2 t^2/2)$**

$$\begin{aligned}
 M(t) &= E(e^{ty}) \\
 &= \int_{-\infty}^{\infty} \exp(ty) f(y) dy \\
 &= \int_{-\infty}^{\infty} \exp(ty) \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right] dy
 \end{aligned}$$

Let  $z = \left(\frac{y-\mu}{\sigma}\right)$ ;  $dz/dy = 1/\sigma$ .

$$\begin{aligned}
 M(t) &= \int_{-\infty}^{\infty} \exp(ty) \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}z^2\right] \sigma dz \\
 &= \int_{-\infty}^{\infty} \exp(t[z\sigma + \mu]) \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}z^2\right] \sigma dz \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}z^2 + t(z\sigma + \mu)\right] dz \\
 &= \frac{\exp(\mu t)}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(z^2 - 2z\sigma t + \{\sigma t\}^2 - \{\sigma t\}^2)\right] dz \\
 &= \frac{\exp(\mu t + \sigma^2 t^2/2)}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(z^2 - 2z\sigma t + \{\sigma t\}^2 - \{\sigma t\}^2)\right] dz \\
 &= \exp(\mu t + \sigma^2 t^2/2) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(z - \sigma t)^2\right] dz \\
 &= \exp(\mu t + \sigma^2 t^2/2)
 \end{aligned}$$

Recognizing that the term inside the integral is the density of  $N(\sigma t, 1)$ .

(7) Show that a binomial random variable  $R$  with denominator  $m$  and probability  $\pi$  has a cumulant generating function  $K(t) = m \log(1 - \pi + \pi e^t)$ . Find  $\lim K(t)$  as  $m \rightarrow \infty$ ,  $\pi \rightarrow 0$  in a way so that  $m\pi \rightarrow \lambda > 0$ . Show that  $Pr(R = r) \rightarrow \lambda^r e^{-\lambda} / r!$  and hence establish that  $R \xrightarrow{D} \text{Poisson}(\lambda)$ . Using your favorite programming language, provide a numerical illustration of the result.

Start with finding the MGF for a binomial distribution:

$$\begin{aligned} MGF_r(t) &= \sum_0^m \binom{m}{r} \pi^r (1 - \pi)^{m-r} e^{tr} \\ &= \sum_0^m \binom{m}{r} (\pi e^t)^r (1 - \pi)^{m-r} \end{aligned}$$

Using the binomial theorem:

$$(x + y)^n = \sum_0^n \binom{n}{k} x^{n-k} y^k$$

We find

$$MGF_r(t) = (1 - \pi + \pi e^t)^m$$

So

$$\begin{aligned} K(t) &= \log[MGF_r(t)] \\ &= \log[(1 - \pi + \pi e^t)^m] \\ &= m \log(1 - \pi + \pi e^t) \end{aligned}$$

Then, let  $m\pi \rightarrow \lambda$ ,  $\lambda > 0$

$$\begin{aligned} \lim_{\substack{m \rightarrow \infty \\ \pi \rightarrow 0}} K(t) &= \lim_{\substack{m \rightarrow \infty \\ \pi \rightarrow 0}} \log[(1 - \pi + \pi e^t)^m] \\ &= \lim_{\substack{m \rightarrow \infty \\ \pi \rightarrow 0}} \log[\{1 + \lambda/m(e^t - 1)\}^m] \\ &= \lambda(1 - e^t) \end{aligned}$$

Using the hint given with  $x \equiv \lambda(1 - e^t)$ . This corresponds to  $MGF_r(t) = e^{\lambda(1 - e^t)}$ . Now, let's check that this corresponds to the MGF of  $\lambda^r e^{-\lambda} / r!$ , which is the pmf of the Poisson distribution.

$$\begin{aligned} MGF_r(t) &= \sum_{r=0}^{\infty} \frac{\lambda^r e^{-\lambda}}{r!} e^{tr} \\ &= \sum_{r=0}^{\infty} \frac{(\lambda e^t)^r e^{-\lambda}}{r!} \end{aligned}$$

Using  $\sum a^x/x! = e^a$

$$\begin{aligned} MGF_r(t) &= e^{-\lambda} e^{\lambda e^t} \\ &= e^{\lambda(e^t - 1)} \end{aligned}$$

The two MGF match, so  $r \xrightarrow{D} \text{Poisson}(\lambda)$ .

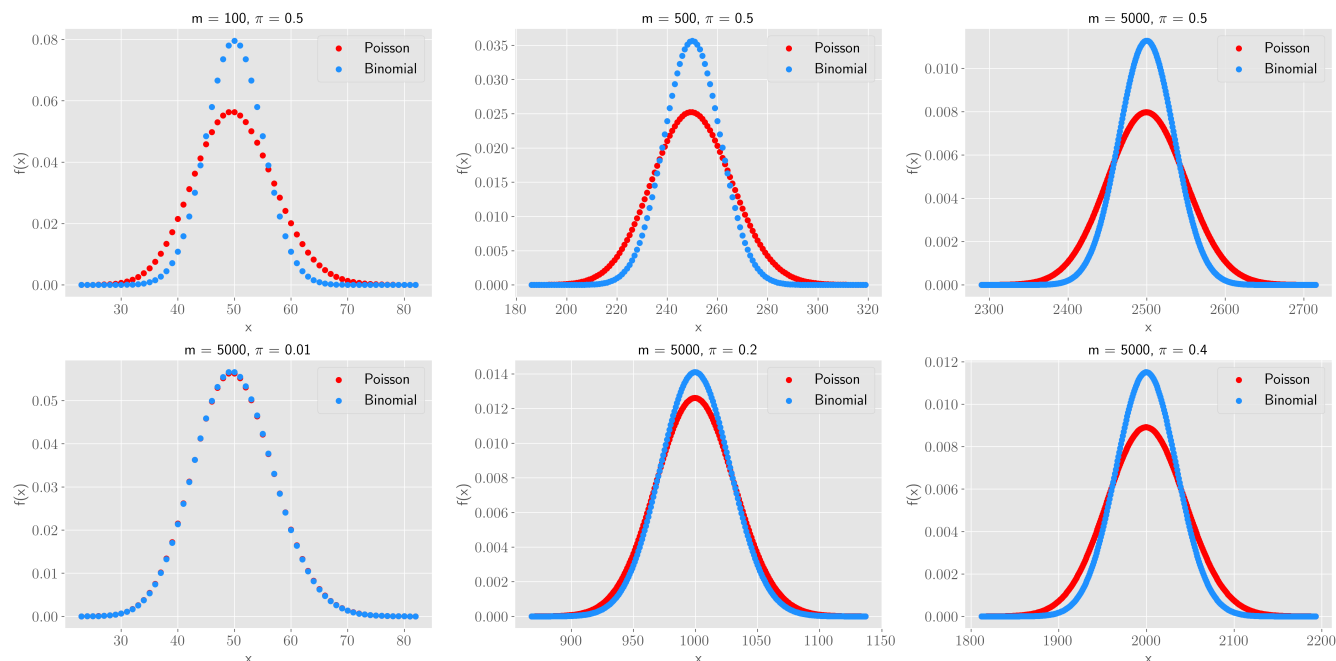


Figure 3: Illustration for binomial becoming Poisson at  $\pi$  becomes small and  $m$  is large.

**(8) If  $Z \sim N(0,1)$ , derive the density of  $Y = Z^2$ . Although  $Y$  is determined by  $Z$ , show that they are uncorrelated.**

We will prove that  $Y \sim \chi_1^2$  by showing that the *MGF* of  $Y$  is the same as that of the  $\chi_1^2$  distribution.

$$\begin{aligned} MGF_y(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} e^{ty} dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(y^2/2)(1-2t)} dy \\ &= (1-2t)^{-1/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}(1-2t)^{-1/2}} e^{-(y^2/2)(1-2t)} dy \\ &= (1-2t)^{-1/2} \end{aligned}$$

By recognizing the pdf of  $N(0, \{1-2t\}^{-1})$ . Now we obtain the MGF of  $\chi_1^2$  using its pdf:



$$\begin{aligned}
MGF_y(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2}\Gamma(1/2)} y^{-1/2} e^{-y/2} e^{ty} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2}\Gamma(1/2)} y^{-1/2} e^{-y/2(1-2t)} dy
\end{aligned}$$

Let  $\tilde{y} = y(1 - 2t)$ , so  $dy/d\tilde{y} = (1 - 2t)^{-1}$

$$\begin{aligned}
MGF_y(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2}\Gamma(1/2)} [\tilde{y}(1 - 2t)^{-1}]^{-1/2} e^{-\tilde{y}/2} (1 - 2t)^{-1} d\tilde{y} \\
&= (1 - 2t)^{-1/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2}\Gamma(1/2)} \tilde{y}^{-1/2} e^{-\tilde{y}/2} d\tilde{y} \\
&= (1 - 2t)^{-1/2}
\end{aligned}$$

From recognizing the pdf of  $\chi_1^2$  which integrates to 1. Since the two MGFs are the same, we have  $Y \sim \chi_1^2$ . Now we show that  $cov(Y, Z) = 0$ .

$$\begin{aligned}
cov(Y, Z) &= E(YZ) - E(Z)E(Y) \\
&= E(Z^3) - E(Z)E(Y) \\
&= 0 - 0 \cdot 1 \\
&= 0
\end{aligned}$$

We use the fact that  $E(Z) = 0$  and that the skewness of  $Z$  is 0; this can be shown using the third derivative of the MGF of  $Z$ .

**(9) Let  $Y = X_1 + bX_2$  where the  $X_j$  are independent normals with means  $\mu_j$  and variances  $\sigma_j^2$ . Show that conditional on  $X_2 = x$ , the distribution of  $Y$  is normal with mean  $\mu_1 + bx$  and variance  $\sigma_1^2$ .**

We start by finding the MGF of  $\mu_1 + bx$ , where  $x$  is a fixed scalar.

$$\begin{aligned}
MFG_y(t) &= E[e^{yt}] \\
&= E[e^{(X_1 + bx)t}] \\
&= e^{bxt} E[e^{X_1 t}]
\end{aligned}$$

Since  $X_1$  is normally distributed with  $\mu_1$  and  $\sigma_1^2$ , and we derived the MGF of the normal distribution earlier, we have:

$$\begin{aligned}
MFG_y(t) &= e^{bxt} e^{\mu_1 t + \sigma_1^2 t^2 / 2} \\
&= e^{(\mu_1 + bx)t + \sigma_1^2 t^2 / 2}
\end{aligned}$$

Which is the MGF of  $\mathcal{N}(\mu_1 + bx, \sigma_1^2)$ . So  $(y|X_2 = x) \sim \mathcal{N}(\mu_1 + bx, \sigma_1^2)$ .  
Then, we show that:

$$\begin{aligned}
& \int \frac{1}{\sigma_1} \phi\left(\frac{y - \mu_1 - bx}{\sigma_1}\right) \frac{1}{\sigma_2} \phi\left(\frac{x - \mu_2}{\sigma_2}\right) dx = \frac{1}{\sqrt{\sigma_1^2 + b^2 \sigma_2^2}} \phi\left(\frac{y - \mu_1 - b\mu_2}{\sigma_1^2 + b^2 \sigma_2^2}\right) \\
& \int \frac{1}{\sigma_1} \phi\left(\frac{y - \mu_1 - bx}{\sigma_1}\right) \frac{1}{\sigma_2} \phi\left(\frac{x - \mu_2}{\sigma_2}\right) dx = \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \frac{[y - \mu_1 - bx]^2}{\sigma_1^2}\right) \left(-\frac{1}{2} \frac{[x - \mu_2]^2}{\sigma_2^2}\right) dx \\
& = \int \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \frac{[y - \mu_1 - bx]^2}{\sigma_1^2}\right) \left(-\frac{1}{2} \frac{[x - \mu_2]^2}{\sigma_2^2}\right) dx \\
& = \int \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \frac{\sigma_2^2[y - \mu_1 - bx]^2 + \sigma_1^2[x - \mu_2]^2}{\sigma_1^2 \sigma_2^2}\right) dx \\
& = \int \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \frac{\sigma_2^2[y - \mu_1 - bx]^2 + \sigma_1^2[x - \mu_2]^2}{\sigma_1^2 \sigma_2^2}\right) dx \\
& = \int \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \frac{\sigma_2^2[y^2 + \mu_1^2 + b^2 x^2 - 2y\mu_1 - 2ybx + 2\mu_1 bx] + \sigma_1^2[x - 2x\mu_2 + \mu_2^2]}{\sigma_1^2 \sigma_2^2}\right) dx \\
& = \int \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \frac{x^2[\sigma_1^2 + b^2 \sigma_2^2] - 2x[\sigma_2^2(by - b\mu_1) + \sigma_1^2 \mu_2^2] + \sigma_2^2[y^2 + \mu_1 - 2y\mu_1] + \sigma_1^2 \mu_2^2}{\sigma_1^2 \sigma_2^2}\right) dx \\
& = \int \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \frac{x^2[\sigma_1^2 + b^2 \sigma_2^2] - 2x[\sigma_2^2(by - b\mu_1) + \sigma_1^2 \mu_2] + \sigma_2^2[y^2 + \mu_1 - 2y\mu_1] + \sigma_1^2 \mu_2^2}{\sigma_1^2 \sigma_2^2}\right) dx
\end{aligned}$$

Let  $\sigma^2 = \sigma_1^2 + b^2 \sigma_2^2$

$$\begin{aligned}
& = \int \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \frac{x^2 - \frac{2x[\sigma_2^2(by - b\mu_1) + \sigma_1^2 \mu_2]}{\sigma^2} + \frac{\sigma_2^2[y^2 + \mu_1 - 2y\mu_1]}{\sigma^2} + \frac{\sigma_1^2 \mu_2^2}{\sigma^2}}{\sigma_1^2 \sigma_2^2 / \sigma^2}\right) dx \\
& = \int \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \frac{\left\{x - \frac{[\sigma_2^2(by - b\mu_1) + \sigma_1^2 \mu_2]}{\sigma^2}\right\}^2 - \left\{\frac{\sigma_2^2(by - b\mu_1) + \sigma_1^2 \mu_2}{\sigma^2}\right\}^2 + \frac{\sigma_2^2[y - \mu_1]^2}{\sigma^2} + \frac{\sigma_1^2 \mu_2^2}{\sigma^2}}{\sigma_1^2 \sigma_2^2 / \sigma^2}\right) dx \\
& = \int \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \frac{\left\{x - \frac{[\sigma_2^2(by - b\mu_1) + \sigma_1^2 \mu_2]}{\sigma^2}\right\}^2}{\sigma_1^2 \sigma_2^2 / \sigma^2}\right) \\
& \quad \exp\left(-\frac{1}{2} \frac{[\sigma_2^2 b(y - \mu_1) + \sigma_1^2 \mu_2]^2 + \sigma^2(\sigma_2^2[y - \mu_1]^2 + \sigma_1^2 \mu_2^2)}{\sigma^2 \sigma_1^2 \sigma_2^2}\right) dx
\end{aligned}$$

Let's focus on the term in the second exponential:

$$\begin{aligned}
& \frac{-[\sigma_2^2 b(y - \mu_1) + \sigma_1^2 \mu_2]^2 + \sigma^2(\sigma_2^2[y - \mu_1]^2 + \sigma_1^2 \mu_2^2)}{\sigma^2 \sigma_1^2 \sigma_2^2} \\
&= \frac{-[\sigma_2^2 b(y - \mu_1) + \sigma_1^2 \mu_2]^2 + (\sigma_1^2 + b^2 \sigma_2^2)(\sigma_2^2[y - \mu_1]^2 + \sigma_1^2 \mu_2^2)}{\sigma^2 \sigma_1^2 \sigma_2^2} \\
&= \frac{-[\sigma_2^4 b^2 (y - \mu_1)^2 + 2\sigma_2^2 b(y - \mu_1)\sigma_1^2 \mu_2 + \sigma_1^4 \mu_2^2] + (\sigma_1^2 + b^2 \sigma_2^2)(\sigma_2^2[y - \mu_1]^2 + \sigma_1^2 \mu_2^2)}{\sigma^2 \sigma_1^2 \sigma_2^2} \\
&= \frac{(y - \mu_1)^2[-\sigma_2^4 b^2 + \sigma_1^2 \sigma_2^2 + b^2 \sigma_2^4] - (y - \mu_1)[2\sigma_1^2 \sigma_2^2 b \mu_2] - \sigma_1^4 \mu_2^2 + \sigma_1^4 \mu_2^2 + b^2 \sigma_1^2 \sigma_2^2 \mu_2^2}{\sigma^2 \sigma_1^2 \sigma_2^2} \\
&= \frac{(y - \mu_1)^2 - 2b\mu_2(y - \mu_1) + b^2 \mu_2^2}{\sigma^2} \\
&= \frac{(y - [\mu_1 + b\mu_2])^2}{\sigma^2}
\end{aligned}$$

Going back and plugging in the term simplified above:

$$\begin{aligned}
&= \int \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \frac{\left\{x - \frac{[\sigma_2^2(by - b\mu_1) + \sigma_1^2 \mu_2]}{\sigma^2}\right\}^2}{\sigma_1^2 \sigma_2^2 / \sigma^2}\right) \exp\left(-\frac{1}{2} \frac{\{y - [\mu_1 + b\mu_2]\}^2}{\sigma^2}\right) dx \\
&= \int \frac{\sigma_1 \sigma_2}{\sigma} \frac{\sigma}{\sigma_1 \sigma_2} \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \frac{\left\{x - \frac{[\sigma_2^2(by - b\mu_1) + \sigma_1^2 \mu_2]}{\sigma^2}\right\}^2}{\sigma_1^2 \sigma_2^2 / \sigma^2}\right) \exp\left(-\frac{1}{2} \frac{\{y - [\mu_1 + b\mu_2]\}^2}{\sigma^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{\{y - [\mu_1 + b\mu_2]\}^2}{\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2 + b\sigma_2^2}} \exp\left(-\frac{1}{2} \frac{\{y - [\mu_1 + b\mu_2]\}^2}{\sigma_1^2 + b\sigma_2^2}\right) \\
&= \frac{1}{\sqrt{\sigma_1^2 + b^2 \sigma_2^2}} \phi\left(\frac{y - \mu_1 - b\mu_2}{\sigma_1^2 + b^2 \sigma_2^2}\right)
\end{aligned}$$

Given that the distribution of the sum of two independent random variables is the convolution of their individual distributions, this gives us the pdf of  $Y$ , i.e.  $Y \sim \mathcal{N}(\mu_1 + b\mu_2, \mu_1^2 + b^2 \mu_2^2)$ .

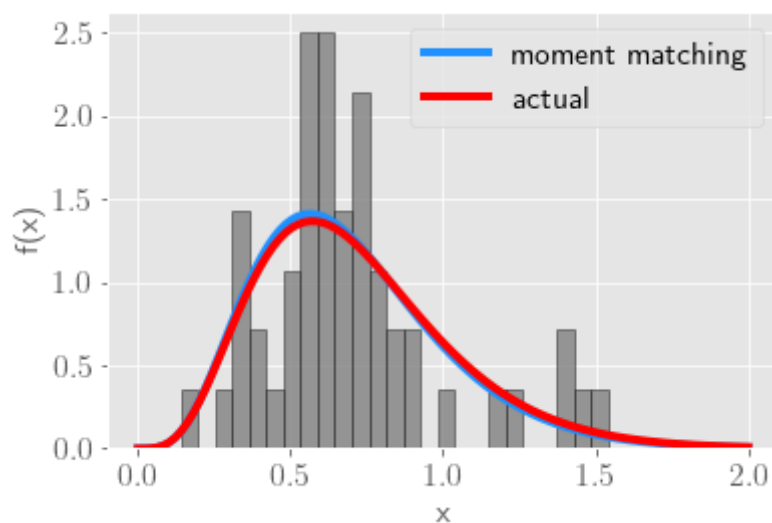
```
In [1]: import matplotlib.pyplot as plt
import numpy as np
from matplotlib import rc
rc('text', usetex=True)
plt.style.use('ggplot')
plt.rcParams.update({'axes.labelsize':16})
plt.rcParams.update({'axes.titlesize':16})
plt.rcParams.update({'legend.fontsize':16})
plt.rcParams['xtick.labelsize'] = 16
plt.rcParams['ytick.labelsize'] = 16
plt.rcParams['lines.linewidth'] = 4
from scipy.stats import gamma, norm, binom, poisson
```

## Problem 5

```
In [2]: alpha, beta = 5, 7
dist = gamma(a=alpha, scale=1/beta)
samples = dist.rvs(50)
```

```
In [3]: mu = samples.mean()
var = samples.var(ddof=1)
beta_guess = mu/var
alpha_guess = mu**2 / var
```

```
In [4]: plt.hist(samples, edgecolor='black', alpha=.8, density=True, color='gray', bins=
x = np.linspace(0, 2, 100)
plt.plot(x, gamma.pdf(x=x, a=alpha_guess, scale=1/beta_guess), color='dodgerblue')
plt.plot(x, dist.pdf(x=x), label='actual', color='red')
plt.xlabel('x')
plt.ylabel('f(x)')
plt.legend()
plt.savefig('first_hist.png', dpi=300, bbox_inches='tight')
```



```
In [5]: def plot(dist, ax, B=500, n=50):
values = []
```

```

for i in range(B):
    samples = dist.rvs(n)
    mu = samples.mean()
    var = samples.var(ddof=1)
    beta_guess = mu/var
    alpha_guess = mu**2 / var
    alpha_hat = mu**2/var
    alpha = dist.mean()*2/ dist.var()
    values.append(np.sqrt(n)* (alpha_hat - alpha) / np.sqrt(2 * alpha_hat *
ax.hist(values, edgecolor='black', alpha=.8, density=True, bins=25, color='g')
x = np.linspace(-3, 3, 100)
ax.plot(x, norm().pdf(x), label='N(0,1)')

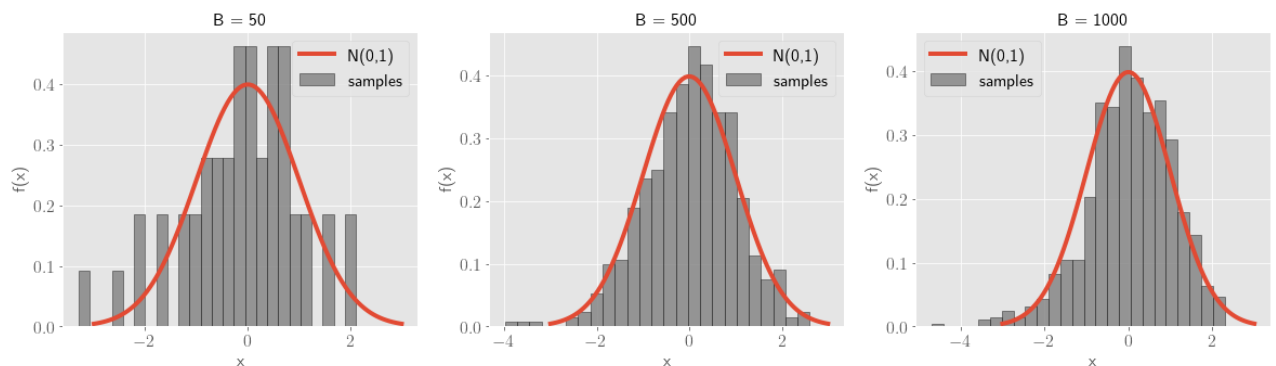
```

In [6]:

```

fig, axes = plt.subplots(figsize=(20, 5), ncols=3)
Bs = [50, 500, 1000]
for B, ax in zip(Bs, axes):
    plot(dist, ax, B=B)
    ax.set_title(f'B = {B}')
    ax.set_xlabel('x')
    ax.set_ylabel('f(x)')
    ax.legend()
plt.savefig('second_hist.png', dpi=300, bbox_inches='tight')

```



## Problem 7

In [7]:

```

def plot(n, p, ax):
    thres = 1e-5
    x = np.arange(*poisson.ppf([thres, 1-thres], mu=n*p), 1)
    ax.scatter(x, poisson.pmf(mu=n*p, k=x), color='red', label='Poisson')
    ax.scatter(x, binom.pmf(k=x, p=p, n=n), color='dodgerblue', label='Binomial')

```

In [8]:

```

fig, axes = plt.subplots(figsize=(20, 10), ncols=3, nrows=2)
Ns = [100, 500, 5000]
for n, ax in zip(Ns, axes[0, :]):
    plot(ax=ax, n=n, p=.5)
    ax.set_title(f'm = {n}, $\\pi$ = 0.5')
    ax.set_xlabel('x')
    ax.set_ylabel('f(x)')
    ax.legend()

Ps = [.01, .2, .4]
for p, ax in zip(Ps, axes[1, :]):
    plot(ax=ax, n=5000, p=p)

```

```

ax.set_title(f'm = 5000, $\pi$ = {p}')
ax.set_xlabel('x')
ax.set_ylabel('f(x)')
ax.legend()
plt.tight_layout()
plt.savefig('third_hist.png', dpi=300, bbox_inches='tight')

```

