The most important questions of life are indeed, for the most part, really only problems of probability. –Pierre-Simon Laplace

Solutions to Problems 10-20

10. $e^{-n}\left(1+\frac{n}{1!}+\frac{n^2}{2!}+\cdots+\frac{n^n}{n!}\right)\xrightarrow{n\to\infty}\frac{1}{2}$

Lemma 1. *The sum of two independent Poisson random variables is a Poisson random variable.* ***Proof of Lemma 1***. *Let $X_1\sim Poi(\lambda_1)$ and $X_2\sim Poi(\lambda_2)$. Then $p_{X_1}(k)=e^{-\lambda_1}\frac{\lambda_1^k}{k!}$, for $k=0,1,2,\ldots$ and $M_{X_1}(t)=\mathbb{E}(e^{tX_1})=\sum_{k=0}^{\infty}e^{tk}e^{-\lambda_1}\frac{\lambda_1^k}{k!}=e^{-\lambda_1}\sum_{k=0}^{\infty}\frac{(\lambda_1e^t)^k}{k!}=e^{\lambda_1(e^t-1)}$. Suppose that $X_1\perp\!\!\!\perp X_2$, then $M_{X_1+X_2}(t)=M_{X_1}(t)M_{X_2}(t)=e^{(\lambda_1+\lambda_2)(e^t-1)}$ which is the MGF of $Poi(\lambda_1+\lambda_2)$.*

Lemma 2. $Poi(n)\xrightarrow{n\to\infty}N(n,n)$. **Proof of Lemma2**. CLT states that for any distribution with finite mean μ and finite variance σ^2 its iid sum tends to $N(\mu,\sigma^2)$ as the number of summands tends to infinity. A corollary of this is that any distribution with finite mean and variance that is an iid sum distribution can be asymptotically approximated by the normal distribution with the same mean and variance as the iid sum. Examples include the binomial distribution (sum of iid Bernoulli), Negative binomial distribution (sum of iid Geometric), Gamma distribution (sum of iid exponential), Chi-square distribution (sum of iid Chi-square), Poisson distribution (sum of iid Poisson), etc. For this case $Poi(n)=\sum_{i=1}^{n}Poi(1)\xrightarrow{n\to\infty}N(n,n)$.

Proof. *Let $Y\sim Poi(n)$, then* $e^{-n}\left(1+\frac{n}{1!}+\frac{n^2}{2!}+\cdots+\frac{n^n}{n!}\right)=\mathbb{P}(Y\le n)\xrightarrow{n\to\infty}\mathbb{P}(N(n,n)\le n)=\frac{1}{2}$.

11. $X_1,X_2,\ldots$ be independent Cauchy r.v.s, each with p.d.f. $f(x)=\frac{d}{\pi(d^2+x^2)}$. Show that $\frac{X_1+X_2+\cdots+X_n}{n}$ has the same distribution as X_1. Does this contradict the WLLT or the CLT?

Answer. The Cauchy distribution does not have an MGF, but it has a CF (characteristic function). We omit the derivation and directly state its form:

$$\varphi_{X_1}(t)=e^{-d|t|}\text{ for }X_1\sim Cauchy(0,d)=\frac{d}{\pi(d^2+x^2)}.$$

For the independent sequence of Cauchy r.v.s $\{X_i\}_{i=1}^n$, the CF of its sample mean is

$$\varphi_{\bar{X}_n}(t)=\prod_{i=1}^{n}\varphi_{X_i}\left(\frac{t}{n}\right)=\prod_{i=1}^{n}e^{-d|t/n|}=e^{-d|t|},$$

the same as $\varphi_X(t)$. This says the distribution of the sample mean of iid Cauchys is the same as any individual Cauchy, which is not anywhere close to a Normal distribution (their CFs are rather different). Does this result contradict the CLT, which implies WLLT? The answer is “No” because Cauchy distribution does *not* possess unique expectation. More specifically, the expectation (uniquely defined for most other distributions) of a Cauchy distribution is everywhere on the real line. The sample space of a Cauchy random variable balances at any point! This is at first very hard to explain because the density clearly has a unique symmetric point at $x = 0$. But after scrutinizing the logic for a while we find a reason: although any symmetry point is an expectation, the converse requires additional conditions to hold, such as an exponentially decreasing probability density or a bounded sample space, etc. The Cauchy distribution does not enjoy any of these regularity conditions. For *any* point x on the real line equipped with Cauchy density, one can always find enough probability mass distributed on extremely remote segments to balance the scale at x. The Cauchy distribution is an extreme case of a heavy-tail distribution.

12. Show that the product of two standard normal random variables which are jointly distributed as the bivariate normal distribution $N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right)$ has moment generating function $(1-2\rho t-(1-\rho^2)t^2)^{-\frac{1}{2}}$. Then deduce its mean and variance.

Solution. Let $X\sim N(0,1)$ and $Y\sim N(0,1)$. Suppose $Corr(X,Y)=\rho$. Then $|\Sigma| = 1-\rho^2$, $\Sigma^{-1}=\frac{1}{1-\rho^2}\begin{bmatrix}1 & -\rho\\ -\rho & 1\end{bmatrix}$ and $f(x,y)=\frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{1}{2}[x\ \ y]\begin{bmatrix}1 & -\rho\\ -\rho & 1\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}}$. Let $Z = XY$. Then

$$M_Z(t)=\mathbb{E}(e^{tXY})=\frac{1}{2\pi\sqrt{1-\rho^2}}=\iint\left\{txy-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\right\}dxdy$$

$$=\frac{1}{2\pi\sqrt{1-\rho^2}}\int\exp\left\{-\frac{\left[x-\left(\rho+t(1-\rho^2)\right)y\right]^2}{2(1-\rho^2)}\right\}dx\int\exp\left\{-\frac{1-\left(\rho+t(1-\rho^2)\right)^2}{2(1-\rho^2)}y^2\right\}dy$$

$$=\frac{1}{2\pi\sqrt{1-\rho^2}}\cdot\sqrt{2\pi(1-\rho^2)}\cdot\sqrt{2\pi\frac{1-\rho^2}{1-\left(\rho+t(1-\rho^2)\right)^2}}=\frac{1}{\sqrt{1-2\rho t-(1-\rho^2)t^2}}$$

The cumulant generating function of Z is the *ln* of MGF:

$$g_Z(t)=\ln M_Z(t)=-\frac{1}{2}\ln(1-2\rho t-(1-\rho^2)t^2)$$

The first two cumulants of Z are the mean and variance, respectively.

$$\mu_Z=g'_Z(0)=\left.\frac{\rho+t(1-\rho^2)}{1-2\rho t-(1-\rho^2)t^2}\right|_{t=0}=\rho$$

$$\sigma_Z^2=g''_Z(0)=\frac{(1-\rho^2)+2\rho^2}{1^2}=1+\rho^2$$

13. The r.v. X_i is normally distributed with mean μ_i and variance σ_i^2, for $i = 1{,}2$, and $X_1 \perp\!\!\!\perp X_2$. Find the distribution of $Z = a_1X_1 + a_2X_2$, where $a_1, a_2 \in \mathbb{R}$.

Solution. Compute the MGF of Z (with $X_1 \perp\!\!\!\perp X_2$ in mind)

$$M_Z(t) = \mathbb{E}\left(e^{t(a_1X_1+a_2X_2)}\right) = \mathbb{E}(e^{ta_1X_1})\mathbb{E}(e^{ta_2X_2}) = e^{a_1\mu_1t+\frac{1}{2}\sigma_1^2(a_1t)^2}e^{a_2\mu_2t+\frac{1}{2}\sigma_2^2(a_2t)^2} = e^{(a_1\mu_1+a_2\mu_1)t+\frac{1}{2}t^2(a_1^2\sigma_1^2+a_2^2\sigma_2^2)}$$

to see that its distribution is just $N(a_1\mu_1 + a_2\mu_1, a_1^2\sigma_1^2 + a_2^2\sigma_2^2)$.

14. Suppose two univariate normal r.v.s $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are also jointly normal with covariance $\begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$. Is the conditional density of $Y|X$ also normal? Derive its form.

Solution. $|\Sigma| = (1-\rho^2)\sigma_X^2\sigma_Y^2$, $\Sigma^{-1} = \frac{1}{(1-\rho^2)\sigma_X^2\sigma_Y^2}\begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}$. Then

$$f(x,y) = \frac{1}{2\pi\sqrt{(1-\rho^2)\sigma_X^2\sigma_Y^2}}e^{-\frac{1}{2(1-\rho^2)\sigma_X^2\sigma_Y^2}[x-\mu_X \quad y-\mu_Y]\begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}\begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix}} = \frac{1}{2\pi\sqrt{(1-\rho^2)\sigma_X^2\sigma_Y^2}}e^{-\frac{1}{2(1-\rho^2)\sigma_X^2\sigma_Y^2}\left((x-\mu_X)^2\sigma_Y^2-2(x-\mu_X)(y-\mu_Y)\rho\sigma_X\sigma_Y+(y-\mu_Y)^2\sigma_Y^2\right)}$$

Then

$$\frac{f(x,y)}{f(x)} = \frac{\sqrt{2\pi\sigma_X^2}}{2\pi\sqrt{(1-\rho^2)\sigma_X^2\sigma_Y^2}}e^{-\frac{1}{2(1-\rho^2)\sigma_X^2\sigma_Y^2}\left((x-\mu_X)^2\sigma_Y^2-2(x-\mu_X)(y-\mu_Y)\rho\sigma_X\sigma_Y+(y-\mu_Y)^2\sigma_Y^2\right)+\frac{1}{2\sigma_X^2}(x-\mu_X)^2}$$

$$= \frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}}e^{-\frac{1}{2(1-\rho^2)\sigma_X^2\sigma_Y^2}\left((x-\mu_X)^2\sigma_Y^2\rho^2-2(x-\mu_X)(y-\mu_Y)\rho\sigma_X\sigma_Y+(y-\mu_Y)^2\sigma_Y^2\right)} = \frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}}e^{-\frac{1}{2(1-\rho^2)\sigma_Y^2}\left((y-\mu_Y)-\frac{\rho\sigma_Y}{\sigma_X}(x-\mu_X)\right)^2}$$

which is just

$$N\left(\mu_Y + (X-\mu_X)\rho\frac{\sigma_Y}{\sigma_X}, \sigma_Y^2(1-\rho^2)\right),$$

a normal distribution with smaller variance.

15. Statisticians are studying relationship between the height of fathers and sons. From the data collected, the average height of the fathers was 1.720 meters; the SD was 0.0696 meters. The average height of the sons was 1.745 meters; the SD was 0.0714 meters. The correlation was 0.501. Assuming normality, True or False and explain: because the sons average 2.5cm taller than the fathers, if the father is 1.830 meters tall, it's 50-50 whether the son is taller than 1.855 meters.

Answer. False. The 50-50 height is $\mu_Y + (X - \mu_X)\rho \frac{\sigma_Y}{\sigma_X}$=1.745+(1.83-1.720)*0.501*0.0714/0.0696=1.80m, 3cm shorter than the given father's height.

16. Let X and Y be two independent standard normal random variables. Find the p.d.f.s of : (i) $X + Y$; (ii) X^2; (iii) $X^2 + Y^2$.

Solution.

(i) From 13 we know that $X + Y \sim N(0,2)$

(ii) X^2 is symmetric about 0. Therefore the density on dx^2 is 2 times the transformed density on dx:

$$\frac{d\mathbb{P}}{d(x^2)} = \frac{d\mathbb{P}}{dx} \cdot \left|\frac{dx}{d(x^2)}\right| \cdot 2 = \frac{1}{\sqrt{2\pi x^2}} e^{-\frac{x^2}{2}},$$

written in terms of the density of $Z = X^2$ is

$$f_Z(z) = \frac{1}{\sqrt{2\pi z}} e^{-\frac{z}{2}} \mathbb{I}(z \geq 0)$$

which is recognized as the chi-square distribution with degree of freedom 1.

(iii) Two independent N(0,1)'s are jointly normal. Thus, $\mathbb{P}(X^2 + Y^2 \leq t) = \frac{1}{2\pi} \iint e^{-\frac{x^2+y^2}{2}} \mathbb{I}(x^2 + y^2 \leq t) dxdy = \frac{1}{2\pi} \iint e^{-\frac{r^2}{2}} \mathbb{I}(r^2 \leq t) r dr d\theta =$

$\int_{r^2=0}^{t} e^{-\frac{r^2}{2}} d\frac{r^2}{2} = \left(1 - e^{-\frac{t}{2}}\right) \mathbb{I}(t \geq 0)$ and the p.d.f. is $f_{X^2+Y^2}(t) = \frac{1}{2} e^{-\frac{t}{2}} \mathbb{I}(t \geq 0)$.

17. Let X and Y be two independent standard normal random variables. Find the joint p.d.f. of $U = X + Y$ and $V = X - Y$. Show that U and V are independent and write down the marginal distribution for U and V.

Solution. Measurability gives the identity

$$f_{X,Y}(x,y)dxdy = f_{U,V}(u,v)dudv.$$

Differential linearity gives the identity

$$\left|\det\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\right| dxdy = dudv$$

or $2dxdy = dudv$. Combining the two to give $f_{U,V}(u,v) = \frac{1}{2}f_{X,Y}(x,y) = \frac{1}{2}\cdot\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}} = \frac{1}{4\pi}e^{-\frac{u^2+v^2}{4}}$. Hence the marginals are $f_U(u) = \frac{1}{\sqrt{4\pi}}e^{-\frac{u^2}{4}}$ and $f_V(v) = \frac{1}{\sqrt{4\pi}}e^{-\frac{v^2}{4}}$. Therefore $U \perp\!\!\!\perp V$ and $U \stackrel{d}{=} V$.

18. Let X and Y be two independent standard normal random variables. Let

$$Z = \begin{cases} |Y|, & \text{if } X > 0,\\ -|Y|, & \text{if } X < 0.\end{cases}$$

By finding $\mathbb{P}(Z \le z)$ for $z < 0$ and $z > 0$, show that $Z \sim N(0,1)$. Explain briefly why the joint distribution of X and Z is not bivariate normal.

Solution. $\mathbb{P}(Z \le z) = \mathbb{P}(|Y| \le z, X > 0) + \mathbb{P}(-|Y| \le z, X < 0) = \frac{1}{2}\mathbb{P}(|Y| \le z) + \frac{1}{2}\mathbb{P}(|Y| \ge -z)$

(i) $z > 0$. Then $\mathbb{P}(|Y| \ge -z) = 1$ and $\mathbb{P}(|Y| \le z) = 2\mathbb{P}(0 \le Y \le z)$. Hence $\mathbb{P}(Z \le z) = \mathbb{P}(0 \le Y \le z) + \frac{1}{2} = \mathbb{P}(Y \le z) = \Phi(z)$

(ii) $z < 0$. Then $\mathbb{P}(|Y| \le z) = 0$ and $\mathbb{P}(|Y| \ge -z) = \mathbb{P}(Y \ge -z \text{ or } Y \le z) = 2\mathbb{P}(Y \le z)$. Hence $\mathbb{P}(Z \le z) = \mathbb{P}(Y \le z) = \Phi(z)$

Therefore $Z \sim N(0,1)$. Since X and Z always have the same sign, they cannot be bivariately normal unless $X = Z$ which is not true.

19. Let $X \sim N(\mu, \sigma^2)$ and suppose $h(x)$ is a smooth bounded function (smooth=any-number-of-times differentiable), $x \in \mathbb{R}$. Prove Stein's formula $\mathbb{E}[(X-\mu)h(X)] = \sigma^2 \mathbb{E}[h'(X)]$.

Solution.

$$\mathbb{E}[(X-\mu)h(X)] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x-\mu)h(x)e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} h(x)d\left(-\sigma^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \sigma^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} h'(x)dx = \sigma^2 \mathbb{E}[h'(X)]$$

which holds because the integrals converge absolutely.

20. Let X be a normally distributed r.v. with mean 0 and variance 1. Compute $\mathbb{E}X^r$ for $r = 0,1,2,3,4$. Let Y be a normally distributed r.v. with mean μ and variance σ^2. Compute $\mathbb{E}Y^r$ for $r = 0,1,2,3,4$.

Solution.

$\mathbb{E}X^0 = 1$. $\mathbb{E}X^1 = \mathbb{E}X^3 = 0$.

$$\mathbb{E}X^2 = \frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx = -\frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^{+\infty} x d(e^{-\frac{x^2}{2}}) = \frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 1$$

$$\mathbb{E}X^4 = \frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^{+\infty} x^4 e^{-\frac{x^2}{2}} dx = -\frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^{+\infty} x^3 d(e^{-\frac{x^2}{2}}) = \frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^{+\infty} 3x^2 e^{-\frac{x^2}{2}} dx = 3$$

$\mathbb{E}Y^0 = 1$.

$\mathbb{E}Y^1 = \mathbb{E}(\sigma X + \mu) = \mu$.

$\mathbb{E}Y^2 = \mathbb{E}(\sigma^2 X^2 + 2\sigma\mu X + \mu^2) = \sigma^2 + \mu^2$.

$\mathbb{E}Y^3 = \mathbb{E}(\sigma^3 X^3 + 3\sigma^2 X^2 \mu + 3\sigma X \mu^2 + \mu^3) = 3\sigma^2\mu + \mu^3$.

$\mathbb{E}Y^4 = \mathbb{E}(\sigma^4 X^4 + 4\sigma^3 X^3 \mu + 6\sigma^2 X^2 \mu^2 + 4\sigma X \mu^3 + \mu^4) = 3\sigma^4 + 6\sigma^2\mu^2 + \mu^4$.