The best way to solve a problem at hand is to look at a harder problem and an easier problem.

Solutions to Problems 1-9

1. *Write down the Law of Iterated Expectations and prove it. Then write down the Law of Total Probability and prove it.*

Law of Iterated Expectations: $(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$. **Proof**. The inner expectation is expanded as $(Y|X) = \int y\mathbb{P}(dy|X) = \int yf(y|X)dy$. Note that this is a transformation on the random variable X. Mount the outer expectation and then expand: $\mathbb{E}\big(\mathbb{E}(Y|X)\big) = \int \mathbb{E}(Y|X)\mathbb{P}(dx) = \int\int yf(y|x)dyf(x)dx = \int\int yf(y|x)f(x)dydx = \int\int yf(x,y)dxdy = \int yf(y)dy = \mathbb{E}(Y)$.

Law of Total Probability (of events): $\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$ where $\{A_i\}_{i=1}^{n}$ is a partition of the sample space S. **Proof**. Let $Y = 1_B$ and $X(\omega) := i$ iff $\omega \in A_i$. Then $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X))$ entails the Law of Total Probability.

2. *Write down Markov's Inequality and prove it. Then write down Chebyshev's Inequality and prove it.*

Markov's Inequality (Only for positive random variable: $X > 0$!): $\mathbb{P}(X \geq k\mathbb{E}(X)) \leq \frac{1}{k}, X > 0$. **Proof**. **Lemma**: For any positive r.v. $X > 0$, $c1_{X\geq c} \leq X$. Thus $1_{X\geq k\mathbb{E}(X)} \leq \frac{X}{k\mathbb{E}(X)}$. Therefore $\mathbb{P}\big(X \geq k\mathbb{E}(X)\big) = \mathbb{E}\big(1_{X\geq k\mathbb{E}(X)}\big) \leq \mathbb{E}\left(\frac{X}{k\mathbb{E}(X)}\right) = \frac{1}{k}$.

Chebyshev's Inequality: $\mathbb{P}\{|X-\mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$, for any $\varepsilon > 0$. **Proof**. $\mathbb{P}\{|X-\mathbb{E}(X)|^2 \geq k\mathbb{V}(X)\} \leq \frac{1}{k}$ ➔ $\mathbb{P}\{|X-\mathbb{E}(X)|^2 \geq \varepsilon^2\} \leq \frac{\mathbb{V}(X)}{\varepsilon^2}$.

3. *Write down the Weak Law of Large Numbers and prove it.*

Weak Law of Large Numbers: Let $\{X_i\}_{i=1}^{n}$ be n i.i.d. random variables with finite variance. Let μ be their common mean. Then $\forall\varepsilon > 0$, $\mathbb{P}\left\{\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| \geq \varepsilon\right\} \xrightarrow{n\to\infty} 0$. **Proof**. From Chebyshev's Inequality, $\mathbb{P}\left\{\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| \geq \varepsilon\right\} \leq \frac{1}{\varepsilon^2}\mathbb{V}\left(\frac{X_1+\cdots+X_n}{n}\right) = \frac{1}{n\varepsilon^2}\mathbb{V}(X_1) \xrightarrow{n\to\infty} 0$.

4. *What is the Moment Generating Function of a distribution? What is the Characteristic Function of a distribution? What are the Moment Generating Function and the Characteristic Function of a univariate normal distribution?*

Moment Generating Function: The moment generating function of the distribution of the r.v. X is a function of some number t: $\mathbb{E}(e^{tX}) = \int e^{tx} f(x)dx = M_X(t)$. Its *r*-th derivative with respect to t at $t = 0$, $\frac{d^r}{dt}\phi(0)$, equals the *r*-th moment $\mathbb{E}(X^r)$.

Characteristic Function: The characteristic function of the distribution of the r.v. X is a function of some number t: $\mathbb{E}\left(e^{itX}\right) = \int e^{itx} f(x)dx = \varphi_X(t)$, where i is the complex unit.

$\varphi_X(t) = M_X(it)$. $M_X(t) = \varphi_X(-it)$.

For the univariate normal distribution $N(\mu, \sigma^2)$, its Moment Generating Function is $e^{\mu t + \frac{1}{2}\sigma^2 t^2}$; its Characteristic function is $e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$.

5. *Write down the density of a univariate normal distribution and show that it integrates to 1.*

Density of the univariate normal distribution: $\phi(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, for $x \in (-\infty, +\infty)$. Next to show it integrates to 1: $\int_{x=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{t=-\infty}^{+\infty} e^{-\frac{t^2}{2}} d(\sigma t + \mu) = \frac{1}{\sqrt{2\pi}} \int_{t=-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \frac{2}{\sqrt{2\pi}} \int_{t=0}^{+\infty} e^{-\frac{t^2}{2}} dt = \frac{2}{\sqrt{2\pi}} \int_{t=0}^{+\infty} e^{-\frac{t^2}{2}} \frac{1}{t} d\frac{t^2}{2} = \frac{2}{\sqrt{2\pi}} \int_{u=0}^{+\infty} e^{-u} \frac{1}{\sqrt{2u}} du = \frac{1}{\sqrt{\pi}} \int_{u=0}^{+\infty} e^{-u} u^{-\frac{1}{2}} du = \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right) = 1$, where $t = \frac{x-\mu}{\sigma}$ and $u = \frac{t^2}{2}$.

6. *Write down the Gamma function $\Gamma(t)$ and show that $\Gamma(n) = (n-1)!$ for any positive integer n.*

Gamma function: $\Gamma(t) = \int_0^\infty e^{-x} x^{t-1} dx$. "Base case": $\Gamma(1) = \int_0^\infty e^{-x} dx = 1$. Next to show the increment relationship "$t\Gamma(t) = \Gamma(t+1)$.": $\Gamma(t) = \int_{x=0}^\infty e^{-x} x^{t-1} dx = \int_{x=0}^{\infty} e^{-x} d\frac{x^t}{t} = \left[\frac{x^t}{t} e^{-x}\right]_{x=0}^{\infty} - \int_{x=0}^{\infty} \frac{x^t}{t} de^{-x} = 0 + \frac{1}{t}\int_{x=0}^\infty e^{-x} x^t dx = \frac{1}{t}\Gamma(t+1)$. Then apply induction principle on the combination of the base case and the increment relationship.

7. *Write down the <u>Central Limit Theorem</u> and prove it.*

<u>Central Limit Theorem:</u> Let $\{X_i\}_{i=1}^{n}$ be n i.i.d. random variables with finite variance. Let μ be their common mean, σ^2 be their common variance. Then $\mathbb{P}\left(\frac{\frac{X_1+\cdots+X_n}{n}-\mu}{\sigma/\sqrt{n}} \le t\right) \xrightarrow{n\to\infty} \Phi(t)$. **Lemma**. $X \perp\!\!\!\perp Y \Longrightarrow \mathbb{E}(XY)=\mathbb{E}(X)\mathbb{E}(Y)$. *<u>Pf of Lemma</u>.* $\mathbb{E}(XY)=\int\int xyf(x,y)dxdy \overset{X\perp\!\!\!\perp Y}{=} \int\int xyf(x)f(y)dxdy = \int xf(x)dx\int yf(y)dy = \mathbb{E}(X)\mathbb{E}(Y)$. **Corollary.** $X \perp\!\!\!\perp Y \Longrightarrow M_{X+Y}(t)=M_X(t)M_Y(t)$. **Proof of CLT**: $\frac{\frac{X_1+\cdots+X_n}{n}-\mu}{\sigma/\sqrt{n}} = \frac{X_1-\mu}{\sigma\sqrt{n}}+\cdots+\frac{X_n-\mu}{\sigma\sqrt{n}}$, $M_{\frac{X_1-\mu}{\sigma\sqrt{n}}+\cdots+\frac{X_n-\mu}{\sigma\sqrt{n}}}(t)=[M_{\frac{X_1-\mu}{\sigma\sqrt{n}}}(t)]^n = [\mathbb{E}(e^{\frac{t}{\sigma\sqrt{n}}(X_1-\mu)})]^n \xrightarrow{n\to\infty} [\mathbb{E}(1+\frac{t}{\sigma\sqrt{n}}(X_1-\mu)+\frac{1}{2}\frac{t^2}{\sigma^2 n}(X_1-\mu)^2)]^n = [(1+\frac{1}{2}\frac{t^2}{\sigma^2 n}\sigma^2)]^n \xrightarrow{n\to\infty} e^{\frac{t^2}{2}}$ which is the MGF of N(0,1). In the proof we have also used classical results of the Euler's number: $\left(1+\frac{x}{n}\right)^n \xrightarrow{n\to\infty} e^x$.

8. *Write down the density of the <u>d-dimensional normal distribution</u>. What are the scalar parameters of a general <u>bivariate normal density</u>? Show that the components of a bivariate normal r.v. are independent of each other iff their <u>correlation coefficient</u> is 0.*

<u>Density of the d-dimensional normal distribution</u>: $\phi(x|\mu,\Sigma)=\frac{1}{(\sqrt{2\pi})^d\sqrt{|\Sigma|}}e^{-\frac{1}{2}(x-\mu)^\intercal\Sigma^{-1}(x-\mu)}$.

<u>For d=2:</u> the vector-matrix parameters are $\mu=\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$, $\Sigma=\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}$, the scalars involved are $\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho$.

Next we show that "Let $\begin{bmatrix}X_1\\ X_2\end{bmatrix}\sim N(\mu,\Sigma)$, then $X_1 \perp\!\!\!\perp X_2$ if and only if $\rho=0$." (if) If $\rho=0$ then $\Sigma=\begin{bmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{bmatrix}$ and then $-\frac{1}{2}(x-\mu)^\intercal\Sigma^{-1}(x-\mu)=-\frac{1}{2}[x_1-\mu_1, x_2-\mu_2]\begin{bmatrix}\frac{1}{\sigma_1^2} & 0\\ 0 & \frac{1}{\sigma_2^2}\end{bmatrix}\begin{bmatrix}x_1-\mu_1\\ x_2-\mu_2\end{bmatrix} = -\frac{1}{2\sigma_1^2}(x_1-\mu_1)^2-\frac{1}{2\sigma_2^2}(x_2-\mu_2)^2$ and then $\frac{1}{(\sqrt{2\pi})^d\sqrt{|\Sigma|}}e^{-\frac{1}{2}(x-\mu)^\intercal\Sigma^{-1}(x-\mu)} = \frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{1}{2}(x_1-\mu_1)^2}\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2}(x_2-\mu_2)^2}$ ➔ $X_1 \perp\!\!\!\perp X_2$.

(only if) $\rho=\frac{\sigma_{12}}{\sigma_1\sigma_2}$, but $\sigma_{12}=\mathbb{E}(X_1-\mu_1)(X_2-\mu_2)=\mathbb{E}(X_1X_2)-\mu_1\mu_2 \overset{X_1\perp\!\!\!\perp X_2}{=} \mathbb{E}(X_1)\mathbb{E}(X_2)-\mu_1\mu_2=0$, therefore $\rho=0$.

9. *Show that the <u>correlation coefficient</u> is absolutely bounded by 1.*

Denote the two r.v.s by X and Y. Consider the non-negative expression $\mathbb{E}(X+kY)^2$: For any k: $0 \leq \mathbb{E}(X+kY)^2 = \mathbb{E}(X^2) + k^2\mathbb{E}(Y^2) + 2k\mathbb{E}(XY)$ ➔ $4[\mathbb{E}(XY)]^2 - 4\mathbb{E}(X^2)\mathbb{E}(Y^2) \leq 0$.