Probability and Statistics for JMC
Solutions 8 — Convergence Concepts

1. Suppose that X is a continuous random variable with pdf
$$f_X(x) = \exp\Big[-(x+2)\Big], \quad \text{for} \quad -2 < x < \infty.$$
Find the mgf of X and then use it to find the expectation and variance of X.

   The mgf is defined as $M(t) = \mathrm{E}(e^{tX})$.
$$\begin{aligned} M(t) = \mathrm{E}(e^{tX}) &= \int_{-\infty}^{\infty} e^{tx} f_X(x)dx = \int_{-2}^{\infty} e^{tx}e^{-(x+2)}dx = \int_{-2}^{\infty} e^{(t-1)x-2}dx \\ &= \frac{1}{t-1}e^{(t-1)x-2}\Big|_{x=-2}^{\infty}. \end{aligned}$$
   We see the integral is only going to converge if $t < 1$. This is fine as we only need the mgf to exist for t in some neighborhood of $t = 0$. For $t < 1$ the mgf is $M(t) = -\frac{1}{t-1}e^{-2t}$.
   The expectation $\mathrm{E}(X)$ is the first moment of X and the variance is related to the second moment $\mathrm{E}(X^2)$.
$$\mathrm{E}(X) = \frac{d}{dt}M(t)\Big|_{t=0} = \left(\frac{1}{(t-1)^2} + \frac{2}{t-1}\right)e^{-2t}\Big|_{t=0} = (1-2)\,e^{-2(0)} = -1.$$
$$\mathrm{E}(X^2) = \frac{d^2}{dt^2}M(t)\Big|_{t=0} = \left(-\frac{2}{(t-1)^3} - \frac{2}{(t-1)^2} - 2\left(\frac{1}{(t-1)^2} + \frac{2}{t-1}\right)\right)e^{-2t}\Big|_{t=0} = 2.$$
   Then $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = 2 - (-1)^2 = 1$.

   To check this, realize that $X = Y - 2$, where $Y \sim$ Exponential(1), so Y has a mean and variance of 1. Then $\mathrm{E}(X) = \mathrm{E}(Y-2) = 1 - 2 = -1$ and $\mathrm{Var}(X) = \mathrm{Var}(Y-2) = \mathrm{Var}(Y) = 1$.

2. Using the Central Limit Theorem, construct normal approximations to each of the following random variables,

   (a) a Binomial distribution $X \sim \mathrm{Binomial}(n, \theta)$;

   Central limit theorem says the "standardized" sample mean of iid $X_1, \dots X_n$, $(\overline{X} - \mu)/(\sigma/\sqrt{n})$, converges in distribution to a standard normal (μ and σ^2 are the mean and variance of one of the X_i's).

   The distribution of a Binomial(n, θ) RV X is exactly that of the sum of n independent Bernoulli(θ) RVs Y_i. So apply the CLT to the Y_i's, noticing that $\overline{Y} = (Y_1 + \cdots + Y_n)/n = X/n$.
   We have $\mathrm{E}(Y_i) = \theta$ and $\mathrm{Var}(Y_i) = \theta(1-\theta)$. The CLT implies that
$$\frac{\overline{Y} - \theta}{\sqrt{\theta(1-\theta)/n}} \xrightarrow{\mathcal{D}} \mathrm{N}(0,1).$$

We rewrite the lefthand side in terms of X, $\dfrac{X/n-\theta}{\sqrt{\theta(1-\theta)/n}}$, and rearrange to get

$$\frac{X-n\theta}{\sqrt{n\theta(1-\theta)}} \xrightarrow{\mathcal{D}} \mathrm{N}(0,1).$$

This shows that, in the limit of large n, X is just a linear transformation of a standard normal Z, i.e. $X = \sqrt{n\theta(1-\theta)}Z + n\theta$. Therefore, X is approximately distributed as a $\mathrm{N}(n\theta,\, n\theta(1-\theta))$. This makes sense because this normal has the same mean and variance as a $\mathrm{Binomial}(n, \theta)$.

(b) a Poisson distribution $X \sim \mathrm{Poisson}(\lambda)$.

2 ways:

- A $\mathrm{Poisson}(\lambda)$ itself is just a $\mathrm{Binomial}(n, \theta)$ RV in the limit where $n \to \infty$, $\theta \to 0$, while $n\theta$ is held constant at the value λ.
  Therefore, for large enough n a $\mathrm{Poisson}(\lambda)$ distribution can be approximated by $\lim_{n\to\infty,\theta\to 0, n\theta=\lambda} \mathrm{N}(n\theta,\, n\theta(1-\theta)) = \mathrm{N}(\lambda, \lambda)$.
- Or realize that $X \sim \mathrm{Poisson}(\lambda)$ is the sum of n independent Y_i, where $Y_i \sim \mathrm{Poisson}(\lambda/n)$ (because the sum of Poisson RVs is also Poisson as we proved directly, and can also be seen taking the product of $\mathrm{Poisson}(\lambda/n)$ mgfs).

  The mean and variance of a Y_i are both equal to λ/n. So the CLT says that the "standardized" sample mean of the Y_i's, $\dfrac{\overline{Y}-\lambda/n}{\sqrt{\lambda}/n}$ is approximately $\mathrm{N}(0, 1)$. Setting $X = n\overline{Y}$ and rearranging gives $X = \sqrt{\lambda}Z + \lambda$, where Z is a standard normal. Thus $X \xrightarrow{\mathcal{D}} \mathrm{N}(\lambda, \lambda)$.

3. Show that for any random variable, X with mean μ and variance σ^2,

$$\mathrm{P}\Big(|X-\mu| \geq t\Big) \leq \frac{\sigma^2}{t^2}.$$

2 ways:

- For X continuous, the lefthand side is the integral of the pdf from $-\infty$ to $\mu - t$ plus the integral from $\mu + t$ to ∞. We know that σ^2 is involved in the answer, so we start with the definition of variance and see if we can manipulate it to get the desired form.

$$\begin{aligned}
\sigma^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \geq \int_{-\infty}^{\mu-t} (x-\mu)^2 f(x)dx + \int_{\mu+t}^{\infty} (x-\mu)^2 f(x)dx \\
&\geq \int_{-\infty}^{\mu-t} t^2 f(x)dx + \int_{\mu+t}^{\infty} t^2 f(x)dx \qquad \text{(since } |x-\mu| \geq t \text{ in both ranges of integration)} \\
&= t^2 \left(\int_{-\infty}^{\mu-t} f(x)dx + \int_{\mu+t}^{\infty} f(x)dx \right) = t^2\, \mathrm{P}\Big(|X-\mu| \geq t\Big).
\end{aligned}$$

If X is discrete the proof goes exactly the same way, just replace the integrals with sums over outcomes of X and the pdf with the pmf. For instance,

$$\int_{-\infty}^{\mu-t} \longrightarrow \sum_{x\in\mathbb{X}:\, x\leq\mu-t}.$$

- Use Chebychev's Inequality, $\mathrm{P}(g(X) \geq r) \leq \dfrac{\mathrm{E}(g(X))}{r}$, where $r > 0$ and $g(\cdot)$ is a non-negative function. Set $r = t^2$ and $g(X) = (X - \mu)^2$.

4. Suppose $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$. Let

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

(a) Show that $\overline{X}$ is a consistent estimator of λ.

We must show that $\overline{X}$ converges in probability to λ, i.e. that $\lim_{n\to\infty} \mathrm{P}(|\overline{X} - \lambda| > \epsilon) = 0$ for any $\epsilon > 0$. We can use the inequality from Q3 applied to the RV $\overline{X}$. We need to use the fact that $\mu = \mathrm{E}(\overline{X}) = \lambda$ and $\sigma^2 = \mathrm{Var}(\overline{X}) = \lambda/n$.

$$\mathrm{P}\Big(|\overline{X} - \lambda| \geq t\Big) \leq \frac{\lambda}{nt^2}.$$

Therefore, as $n \to \infty$ the probability that $|\overline{X} - \lambda| > \epsilon$ is zero for any $\epsilon > 0$.

(b) Suppose $T_n = \exp(-\overline{X})$, show that $T_n \xrightarrow{\mathcal{P}} \exp(-\lambda)$.

Notice that $g(x) = \exp(-x)$ is a continuous, 1-to-1 function and that $g(\lambda) = \exp(-\lambda)$. Therefore, the probability that T_n is within ϵ of $\exp(-\lambda)$ is equal to the probability that $\overline{X}$ is within some particular interval surrounding λ. As $n \to \infty$ this latter probability goes to 1 (part (a)), which proves that $T_n \xrightarrow{\mathcal{P}} \exp(-\lambda)$.

Explicitly, $\exp(-\lambda) - \epsilon < T_n < \exp(-\lambda) + \epsilon \iff \lambda - a < \overline{X} < \lambda + b$ for some positive numbers a and b (you can find a and b if you want by writing T_n in terms of $\overline{X}$ and rearranging the first set of inequalities, but it's not necessary). Therefore,

$$\begin{aligned}
\mathrm{P}\Big(|T_n - \exp(-\lambda)| < \epsilon\Big) &= \mathrm{P}\Big(\lambda - a < \overline{X} < \lambda + b\Big) \\
&\geq \mathrm{P}\Big(\lambda - \min(a,b) < \overline{X} < \lambda + \min(a,b)\Big) \\
&= \mathrm{P}\Big(|\overline{X} - \lambda| < \min(a,b)\Big),
\end{aligned}$$

and this last probability goes to 1 as $n \to \infty$.

5. Markov's inequality states that, for any $r > 0$, and for an arbitrary random variable Y,

$$\mathrm{P}(|Y| > \delta) \leq \frac{\mathrm{E}(|Y|^r)}{\delta^r},$$

for all $\delta > 0$. Use this result to show that for independent and identically distributed random variables $X_1, \ldots X_n$ with mean μ and variance $\sigma^2 < \infty$, the sample mean $\overline{X}$ converges in probability to μ as $n \to \infty$.

Set $Y = \overline{X} - \mu$ and $r = 2$. Then the numerator on the righthand side is the variance of the sample mean, which is σ^2/n. So $\mathrm{P}(|\overline{X} - \mu| > \delta) \leq \frac{\sigma^2}{n\delta^2}$. For large n the righthand side goes to 0. Therefore, $\overline{X} \xrightarrow{\mathcal{P}} \mu$. (This is very similar to Q3 + Q4a).

6. Find the moment generating function of the random variable X which has density function

$$f_X(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2, \\ 0 & \text{otherwise.} \end{cases}$$

Write down the moment generating function (MGF) of X, and by expanding the MGF as a power series in t, find the mean and variance of X.

MGF is $M(t) = \mathrm{E}(e^{tX}) = \int_0^2 e^{tx}\frac{1}{2}x\,dx$. Integrate by parts to get

$$\begin{aligned} M(t) &= \frac{1}{2}x\frac{e^{tx}}{t}\Big|_{x=0}^{2} - \int_0^2 \frac{e^{tx}}{2t}dx = \frac{1}{2}x\frac{e^{tx}}{t}\Big|_{x=0}^{2} - \int_0^2 \frac{e^{tx}}{2t}dx = \frac{e^{2t}}{t} - \frac{e^{tx}}{2t^2}\Big|_{x=0}^{2} \\ &= \frac{e^{2t}}{t} - \frac{e^{2t}}{2t^2} + \frac{1}{2t^2}. \end{aligned}$$

When expanded as a power series in t the MGF has coefficients that are the moments of X:

$$\begin{aligned} M(t) = \mathrm{E}(e^{tX}) &= \mathrm{E}\left[1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots\right] \\ &= 1 + \mathrm{E}(X)t + \frac{\mathrm{E}(X^2)}{2!}t^2 + \frac{\mathrm{E}(X^3)}{3!}t^3 + \cdots. \end{aligned}$$

We can find the coefficients for $M(t) = \frac{e^{2t}}{t} - \frac{e^{2t}}{2t^2} + \frac{1}{2t^2}$ either by taking derivatives wrt t and then setting $t = 0$ or by writing the exponentials as power series and collecting terms for each power of t. See Q1 for the first method. Expanding the exponentials we have (we only need to expand enough to get to t^2 since the question only asks for the mean and variance),

$$\begin{aligned} M(t) &= \frac{1 + 2t + \frac{(2t)^2}{2!} + \frac{(2t)^3}{3!} + \cdots}{t} - \frac{1 + 2t + \frac{(2t)^2}{2!} + \frac{(2t)^3}{3!} + \frac{(2t)^4}{4!} + \cdots}{2t^2} + \frac{1}{2t^2} \\ &= \left(\frac{1}{t} + 2 + 2t + \frac{8}{6}t^2 + \cdots\right) - \left(\frac{1}{2t^2} + \frac{1}{t} + 1 + \frac{4}{6}t + \frac{8}{24}t^2 + \cdots\right) + \frac{1}{2t^2} \\ &= \frac{0}{t^2} + \frac{0}{t} + 1 + \left(2 - \frac{4}{6}\right)t + \left(\frac{8}{6} - \frac{8}{24}\right)t^2 + \cdots \\ &= 1 + \frac{4}{3}t + t^2 + \cdots. \end{aligned}$$

Equating each power of t we get $\mathrm{E}(X) = 4/3$ and $\mathrm{E}(X^2)/2! = 1$. Then $\mathrm{Var}(X) = 2 - (4/3)^2 = 2/9$.