MATH 6605 Probability Theory
Section 9.3: Moment Generating Functions (continued)
Section 9.4: Fubini's Theorem

February 27, 2025

Recap of Moment Generating Functions:

Definition 9.3.A. The moment generating function (MGF) of a random variable X is the function $M_X : \mathbb{R} \to [0, \infty]$ defined by

$$M_X(s) \;:=\; E\left(e^{sX}\right) \qquad (s \in \mathbb{R})\,.$$

Properties 9.3.B. (ii) If X and Y are independent random variables, then $M_{X+Y}(s) \,=\, M_X(s)\, M_Y(s)$.

Theorem 9.3.3: *Let X be a random variable. Assume $M_X(s) < \infty$ for every $s \in (-a, a)$, for some $a > 0$. Then*

$$M_X(s) \;=\; \sum_{k=0}^{\infty} \frac{s^k E(X^k)}{k!} \qquad \textit{for all } s \in (-a, a).$$

Consequently,

$$\left.\frac{d^k\, M_X(s)}{ds^k}\right|_{s=0} \;=\; E(X^k).$$

Example: We calculated in class that if X has the exponential distribution with parameter $\beta > 0$ (which we can write briefly as "$X \sim \text{Exponential}(\beta)$"), then

$$M_X(s) \;=\; \begin{cases} \frac{\beta}{\beta - s} & \text{if } s < \beta, \\ \infty & \text{if } s \geq \beta. \end{cases}$$

From this, we can compute the moments $E(X^k)$ of X:

$$M'(s) \;=\; \frac{\beta}{(\beta - s)^2} \qquad \therefore\; E(X) \;=\; M'(0) \;=\; \frac{\beta}{(\beta - 0)^2} \;=\; \frac{1}{\beta}.$$

$$M''(s) \;=\; \frac{2\beta}{(\beta - s)^3} \qquad \therefore\; E(X^2) \;=\; M''(0) \;=\; \frac{2\beta}{(\beta - 0)^3} \;=\; \frac{2}{\beta^2}.$$

$$\therefore\; \text{Var}(X) \;=\; E(X^2) - (E(X))^2 \;=\; \frac{2}{\beta^2} - \left(\frac{1}{\beta}\right)^2 \;=\; \frac{1}{\beta^2}.$$

And so on.

Alternatively, we can compute the moments using the power series expansion:

$$M_X(s) \;=\; \frac{1}{1-\frac{s}{\beta}} \;=\; \sum_{k=0}^{\infty} \frac{s^k}{\beta^k} \qquad \text{for } -\beta < s < \beta.$$

Recalling from Theorem 9.3.3 that $\quad M_X(s) \;=\; \displaystyle\sum_{k=0}^{\infty} \frac{s^k E(X^k)}{k!}$,

we observe $\quad \dfrac{E(X^k)}{k!} \;=\; \dfrac{1}{\beta^k} \quad$ for every k; $\quad \therefore\; E(X^k) \;=\; \dfrac{k!}{\beta^k}$.

Application of MGFs: Large Deviations

We shall illustrate through the following detailed example.

Let H_n represent the number of Heads in n tosses of a fair coin. In our probability triple $(\Omega_*, \sigma_E, \mathbf{P}_E)$ of sequences of tosses of a fair coin, we take $H_n(\omega)$ to be the number of H's in $(\omega_1, \ldots, \omega_n)$. We know that H_n has the binomial distribution with parameters n and $1/2$, and $E(H_n) = n/2$.
We know that H_n/n converges to $\frac{1}{2}$ in probability (and almost surely) by the laws of large numbers. In particular, we know that $\mathbf{P}(H_n \geq 0.6n)$ goes to 0 as $n \to \infty$, but how quickly?
Recall how we used Chebyshev's Inequality to prove the WLLN:

$$\begin{aligned}
\mathbf{P}\left(H_n \geq \left(\frac{1}{2} + 0.1\right) n\right) &= \mathbf{P}\left(\frac{H_n}{n} - \frac{1}{2} \geq 0.1\right) \\
&\leq \mathbf{P}\left(\left|\frac{H_n}{n} - \frac{1}{2}\right| \geq 0.1\right) \\
&\leq \frac{\operatorname{Var}\left(\frac{H_n}{n}\right)}{(0.1)^2} = \frac{\text{constant}}{n}.
\end{aligned}$$

But in fact, the probability decays much faster, as we now show.

Theorem 9.3.C: (Special case of Theorem 9.3.4) *Let H_n have the binomial distribution with parameters n and $1/2$ (corresponding to the number of Heads in n tosses of a fair coin; note that $E(H_n/n) = 1/2$). Assume $1/2 < b < 1$. Then*

$$\mathbf{P}\left(\frac{H_n}{n} \geq b\right) \leq \rho^n, \quad \textit{where } \rho = \frac{1}{2\,(1-b)^{1-b}\, b^b} < 1.$$

To illustrate, for $b = 0.6$, we get $\rho = 0.980\ldots$. Then $\rho^{100} = 0.133\ldots$, and $\rho^{1000} \approx 1.7 \times 10^{-9}$.

Proof of Theorem 9.3.C: We follow the text's proof of 9.3.4.
Fix $b > 1/2$ (and $b < 1$).
Let $X_1, X_2, \ldots$ be i.i.d. random variables with
$\mathbf{P}(X_i = 0) = \frac{1}{2} = \mathbf{P}(X_i = 1)$. Then $\mathcal{L}(H_n) = \mathcal{L}(X_1 + \cdots + X_n)$.

$\frac{1}{2} < b < 1$; $X_1, X_2, \ldots$ are i.i.d. with $\mathbf{P}(X_i = 0) = \frac{1}{2} = \mathbf{P}(X_i = 1)$; $\mathcal{L}(H_n) = \mathcal{L}(X_1 + \cdots + X_n)$.

$$
\begin{aligned}
\mathbf{P}(H_n \geq bn) &= \mathbf{P}\left(\sum_{i=1}^{n} X_i \, - \, bn \, \geq \, 0\right) \\
&= \mathbf{P}\left(\sum_{i=1}^{n} (X_i - b) \, \geq \, 0\right) \\
&\leq \mathbf{P}\left(e^{s \sum_{i=1}^{n} (X_i - b)} \geq \, e^0\right) \qquad \text{for all } s > 0 \\
&\leq E\left(e^{s \sum_{i=1}^{n} (X_i - b)}\right) / e^0 \qquad \text{(by Markov's Inequality)} \\
&= E\left(e^{s(X_1 - b)} e^{s(X_2 - b)} \cdots e^{s(X_n - b)}\right) \\
&= E\left(e^{s(X_1 - b)}\right) \, E\left(e^{s(X_2 - b)}\right) \cdots E\left(e^{s(X_n - b)}\right) \\
&= \left(E\left(e^{s(X_1 - b)}\right)\right)^n \, = \, (M_{X_1 - b}(s))^n \qquad (\{X_i\} \text{ i.i.d}).
\end{aligned}
$$

$\frac{1}{2} < b < 1$; $X_1, X_2, \ldots$ are i.i.d. with $\mathbf{P}(X_i = 0) = \frac{1}{2} = \mathbf{P}(X_i = 1)$;
$\mathcal{L}(H_n) = \mathcal{L}(X_1 + \cdots + X_n)$.
(*) $\quad \mathbf{P}(H_n \geq bn) \leq (M_{X_1 - b}(s))^n \qquad$ for all $s > 0$.

$$\text{Now,} \quad M_{X_1-b}(s) = E\left(e^{s(X_1-b)}\right) = \frac{1}{2}e^{s(0-b)} + \frac{1}{2}e^{s(1-b)}.$$

$$M'_{X_1-b}(s) = \frac{1}{2}\left(-be^{-sb} + (1-b)e^{s(1-b)}\right)$$

$$\therefore\ M'_{X_1-b}(0) = \frac{1}{2}(1-2b) < 0 \quad (= E(X_1 - b) < 0)$$

Also note $M''_{X_1-b}(s) > 0$. So we can find the best s for (*) [i.e., minimize $M_{X_1-b}(s)$] by solving $M'(s) = 0$:

$$be^{-sb} = (1-b)e^{s(1-b)} \quad \Rightarrow \quad e^s = \frac{b}{1-b} \quad \therefore \text{Let } s^* = \ln\left(\frac{b}{1-b}\right).$$

Notice $s^* > 0$ since $b > \frac{1}{2}$.
Now substitute $s = s^*$ into (*): $\quad \mathbf{P}(H_n \geq bn) \leq (M_{X_1-b}(s^*))^n$

$$\begin{aligned}
M_{X_{1-b}}(s^*) &= \frac{1}{2}\left(e^{-s^* b} + e^{s^*(1-b)}\right) \\
&= \frac{1}{2}\left(e^{s^*}\right)^{-b}(1+e^{s^*}) \qquad \text{Recall } e^{s^*} = \frac{b}{1-b}: \\
&= \frac{1}{2}\left(\frac{b}{1-b}\right)^{-b}\left(1+\frac{b}{1-b}\right) \qquad \text{Do some algebra...} \\
&= \frac{1}{2\,b^b\,(1-b)^{1-b}} \quad < M_X(0) = 1. \\
\text{Let } \rho &= \frac{1}{2\,b^b\,(1-b)^{1-b}}.
\end{aligned}$$

Then we have $\mathbf{P}(H_n \geq bn) \leq (M_{X_{1-b}}(s^*))^n = \rho^n$. Q.E.D.

Remark: It can in fact be proved that $\lim_{n\to\infty} \mathbf{P}(H_n \geq bn)^{1/n} = \rho$.

Section 9.4: Fubini's Theorem

Recall *product measure* from Section 2.7:
Let $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$ be probability triples. Their *product measure* triple is the probability triple $(\Omega_1 \times \Omega_2, \mathcal{F}_{12}, \mathbf{P}_{12})$ characterized by $\mathbf{P}_{12}(A \times B) = \mathbf{P}_1(A)\,\mathbf{P}_2(B)$ for $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$.
(We can also write this $\mathbf{P}_{12}$ as $\mathbf{P}_1 \times \mathbf{P}_2$.)

For random variables X on the product triple, we can evaluate $E(X)$ as usual (start with simple functions on $\Omega_1 \times \Omega_2$, etc.), but we can also evaluate it by iterated expectations involving only one of $\mathbf{P}_1$ or $\mathbf{P}_2$ at a time.

Notation: For a random variable X on $(\Omega, \mathcal{F}, \mathbf{P})$, we can write the expectation as a Lebesgue-type integral:

$$E_{\mathbf{P}}(X) = \int_\Omega X(\omega)\,\mathbf{P}(d\omega) = \int_\Omega X\,d\mathbf{P}\,.$$

For the product measure, this is

$$E_{\mathbf{P}_1 \times \mathbf{P}_2}(X) = \int_{\Omega_1 \times \Omega_2} X(\omega_1, \omega_2)\,d(\mathbf{P}_1 \times \mathbf{P}_2)\,.$$

Theorem 9.4.1: Fubini's Theorem. *For the product measures on the previous slide, let X be a random variable on $(\Omega_1 \times \Omega_2, \mathcal{F}_{12}, \mathbf{P}_1 \times \mathbf{P}_2)$ whose expected value is defined. Then*

$$\begin{aligned}
E_{\mathbf{P}_1 \times \mathbf{P}_2}(X) &= \int_{\Omega_1 \times \Omega_2} X(\omega_1, \omega_2)\, d(\mathbf{P}_1 \times \mathbf{P}_2) \\
&= \int_{\Omega_1} \left(\int_{\Omega_2} X(\omega_1, \omega_2)\, \mathbf{P}_2(d\omega_2) \right) \mathbf{P}_1(d\omega_1) \qquad (1) \\
&= \int_{\Omega_2} \left(\int_{\Omega_1} X(\omega_1, \omega_2)\, \mathbf{P}_1(d\omega_1) \right) \mathbf{P}_2(d\omega_2) \,. \qquad (2)
\end{aligned}$$

(Proof omitted.)

We can also write (1) as $E_{\mathbf{P}_1}\left(E_{\mathbf{P}_2} X(\omega_1, \omega_2)\right)$, where the inner quantity $E_{\mathbf{P}_2} X(\omega_1, \omega_2)$ is interpreted as having ω_1 fixed, and the expected value taken with respect to $\omega_2 \in \Omega_2$.

Thus, $E_{\mathbf{P}_2} X(\omega_1, \omega_2)$ is a function of ω_1, which is measurable with respect to $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$ (measurability can be proved).

Important case: Suppose $X = g(Y_1(\omega_1), Y_2(\omega_2))$, where each Y_i is a random variable on $(\Omega_i, \mathcal{F}_i, \mathbf{P}_i)$. By construction of product measure, Y_1 and Y_2 are independent, because

$$\begin{aligned}
\mathbf{P}_{12}(Y_1(\omega_1)\in A, Y_2(\omega_2)\in B) &= \mathbf{P}_{12}\left((\omega_1,\omega_2) \in Y_1^{-1}(A)\times Y_2^{-1}(B)\right) \\
&= \mathbf{P}_1(Y_1^{-1}(A))\,\mathbf{P}_2(Y_2^{-1}(B)) \\
&= \mathbf{P}_1(Y_1 \in A))\,\mathbf{P}_2(Y_2 \in B).
\end{aligned}$$

Then $E_{\mathbf{P}_2}\left(g(Y_1(\omega_1), Y_2(\omega_2))\right) \;=\; g_1(Y_1(\omega_1))$ where the function g_1 on $\mathbb{R}$ is defined by $g_1(y_1) = E_{\mathbf{P}_2}(g(y_1, Y_2))$. Then Fubini's Theorem says that

$$E_{\mathbf{P}_{12}}(g(Y_1, Y_2)) \;=\; E_{\mathbf{P}_1}(g_1(Y_1)).$$

To illustrate, we shall consider the important example of computing the distribution of the sum of two independent random variables.

Let Y and Z be independent random variables, with c.d.f.'s F_Y and F_Z. We want to compute the c.d.f. of the random variable $Y + Z$.

We take $\mathbf{P}_1$ to be μ_Y, and $\mathbf{P}_2$ to be μ_Z.
Correspondingly, $\Omega_1 = \Omega_2 = \mathbb{R}$ and $\mathcal{F}_1 = \mathcal{F}_2 = \mathcal{B}(\mathbb{R})$.
Write $\mathbf{P}_{YZ}$ for $\mu_Y \times \mu_Z$, and (y, z) for (ω_1, ω_2).

Fix $t \in \mathbb{R}$, and let $g(y,z) = \mathbf{1}_{\{y+z \le t\}}(y,z)$. Then

$$\begin{aligned}
\mathbf{P}_{YZ}(Y + Z \le t) &= E_{\mathbf{P}_{YZ}} g(Y, Z) \\
&= E_{\mu_Y}(g_2(Y)) \\
\text{where} \quad g_2(y) &= E_{\mu_Z}(g(y, Z)) \\
&= \mu_Z(y + Z \le t) \\
&= \mu_Z(Z \le t - y) \qquad = F_Z(t - y). \\
\therefore\ \mathbf{P}_{YZ}(Y + Z \le t) &= E_{\mu_Y}\left(F_Z(t - Y)\right) \\
&= \int_{\mathbb{R}} F_Z(t - y)\, \mu_Y(dy).
\end{aligned}$$

See Theorem 9.4.5 for a more general statement.

$$\mathbf{P}_{YZ}(Y+Z \le t) \;=\; E_{\mu_Y}\left(F_Z(t-Y)\right) \;=\; \int_{\mathbb{R}} F_Z(t-y)\,\mu_Y(dy).$$

If Y has pdf f_Y, then this becomes

$$\mathbf{P}_{YZ}(Y+Z \le t) \;=\; \int_{-\infty}^{\infty} F_Z(t-y)\, f_Y(y)\, dy\,.$$

If also Z has pdf f_Z, then we have

$$\begin{aligned} F_Z(t-y) \;&=\; \int_{-\infty}^{t-y} f_Z(z)\, dz \qquad \text{Substitute } z = u-y,\, dz = du \\ &=\; \int_{-\infty}^{t} f_Z(u-y)\, du\,, \\ \therefore\quad \mathbf{P}_{YZ}(Y+Z \le t) \;&=\; \int_{-\infty}^{\infty}\int_{-\infty}^{t} f_Z(u-y)\, du\, f_Y(y)\, dy \\ &=\; \int_{-\infty}^{t}\int_{-\infty}^{\infty} f_Z(u-y)\, f_Y(y)\, dy\, du \end{aligned}$$

Hence $Y+Z$ has pdf $\quad f_{Y+Z}(u) \;=\; \int_{-\infty}^{\infty} f_Z(u-y)\, f_Y(y)\, dy.$

The integral expression $f_{Y+Z}(u) = \int_{-\infty}^{\infty} f_Z(u-y)\, f_Y(y)\, dy$ is called the convolution of f_Y and f_Z.

There is a similar formula in the discrete case, which is easier to derive. Suppose Y and Z are independent integer-valued random variables. Then for each integer k, we have

$$\begin{aligned}
\mathbf{P}(Y+Z=k) &= \sum_{i,j:\, i+j=k} \mathbf{P}(Y=i,\, Z=j) \\
&= \sum_{i,j:\, i+j=k} \mathbf{P}(Y=i)\, \mathbf{P}(Z=j) \\
&= \sum_{i\in\mathbb{Z}} \mathbf{P}(Y=i)\, \mathbf{P}(Z=k-i).
\end{aligned}$$

<u>Exercise 9.5.15:</u> Use the above formula to show that the sum of two independent Poisson random variables has a Poisson distribution.