

Lecture 8: Sums of Independent Random Variables and Moment Generating Functions

EECS 126 (UC Berkeley)

Spring 2026 — February 12, 2026

Reading: Bertsekas & Tsitsiklis, §4.4–4.6.

These are my personal notes from Professor Kannan Ramchandran's EECS 126 (Probability and Random Processes) course at UC Berkeley.

Topics: sums of independent random variables (convolution), tail bounds, and moment generating functions (MGFs).

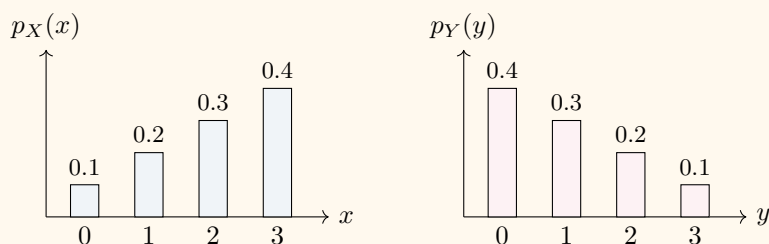
8.1 Discrete Convolution

Example 8.1 (Sum of two independent discrete random variables)

Let X and Y be independent, each taking values in $\{0, 1, 2, 3\}$, with

$$p_X = (0.1, 0.2, 0.3, 0.4), \quad p_Y = (0.4, 0.3, 0.2, 0.1) \quad (8.1)$$

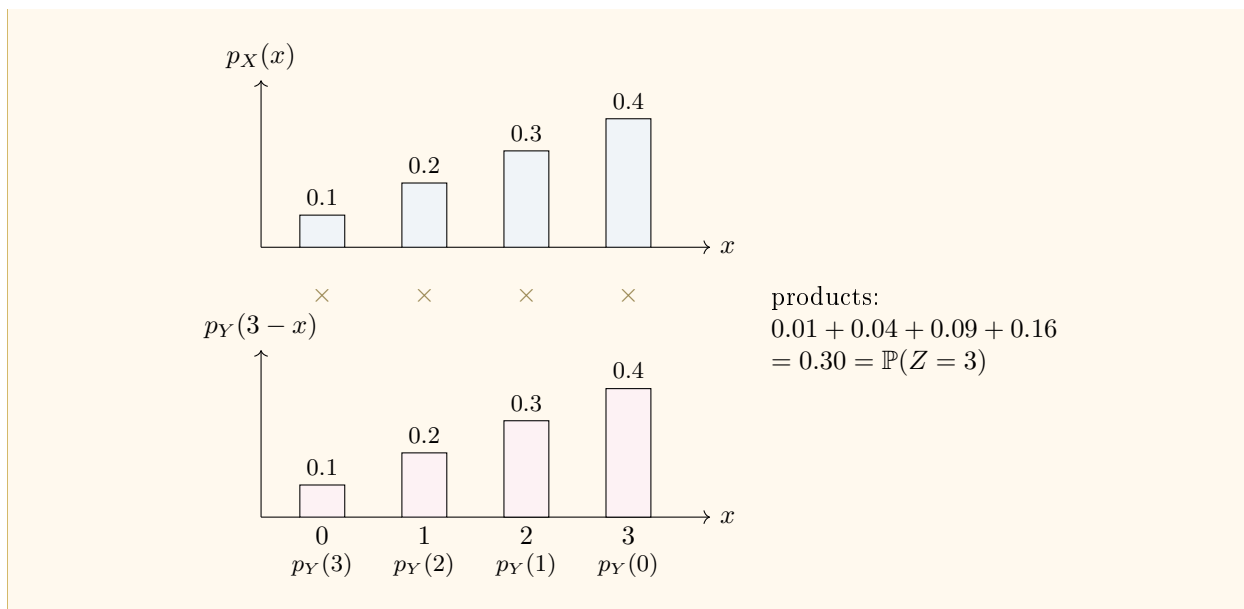
at the values 0, 1, 2, 3 respectively. Let $Z = X + Y$. Find $\mathbb{P}(Z = 3)$.



$Z = 3$ occurs exactly for the pairs $(x, y) \in \{(0, 3), (1, 2), (2, 1), (3, 0)\}$, i.e. $y = 3 - x$:

$$\begin{aligned} \mathbb{P}(Z = 3) &= \mathbb{P}(X + Y = 3) = \sum_{(x,y): x+y=3} \mathbb{P}(X = x, Y = y) \\ &= \sum_{(x,y): x+y=3} \mathbb{P}(X = x) \mathbb{P}(Y = y) \quad (\text{independence}) \\ &= \sum_x \mathbb{P}(X = x) \mathbb{P}(Y = 3 - x) \\ &= p_X(0) p_Y(3) + p_X(1) p_Y(2) + p_X(2) p_Y(1) + p_X(3) p_Y(0) \\ &= (0.1)(0.1) + (0.2)(0.2) + (0.3)(0.3) + (0.4)(0.4) \\ &= 0.01 + 0.04 + 0.09 + 0.16 = 0.30. \end{aligned} \quad (8.2)$$

To see the products line up, flip p_Y and shift it by 3: plot $p_Y(3 - x)$ against x , directly under $p_X(x)$, then multiply bar by bar and add.



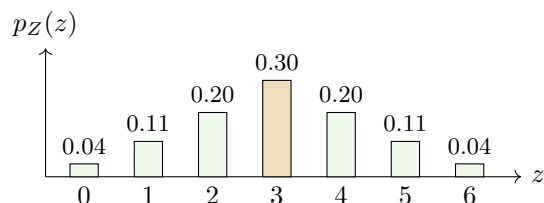
Definition 8.1 (Discrete convolution)

If X and Y are independent and integer-valued, then $Z = X + Y$ has PMF

$$p_Z(z) = (p_X * p_Y)(z) \triangleq \sum_x p_X(x) p_Y(z - x). \quad (8.3)$$

Here x is a dummy summation variable.

Applying (8.3) at every z in the example gives the full PMF of Z :



Check: $0.04 + 0.11 + 0.20 + 0.30 + 0.20 + 0.11 + 0.04 = 1$.

8.2 Continuous Convolution

Theorem 8.1 (Density of a sum of independent random variables)

If X and Y are independent continuous random variables and $Z = X + Y$, then

$$f_Z(z) = (f_X * f_Y)(z) \triangleq \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx. \quad (8.4)$$

Here x is a dummy variable of integration.

Proof by conditioning. Find the conditional CDF of Z given $X = x$, then differentiate:

$$\begin{aligned}
 F_{Z|X}(z | x) &= \mathbb{P}(Z \leq z | X = x) = \mathbb{P}(X + Y \leq z | X = x) \\
 &= \mathbb{P}(x + Y \leq z | X = x) \\
 &= \mathbb{P}(Y \leq z - x | X = x) \\
 &= \mathbb{P}(Y \leq z - x) && \text{(independence)} \\
 &= F_Y(z - x).
 \end{aligned} \tag{8.5}$$

Differentiating in z ,

$$f_{Z|X}(z | x) = \frac{d}{dz} F_Y(z - x) = f_Y(z - x). \tag{8.6}$$

Multiply both sides by the marginal $f_X(x)$; the left side becomes the joint PDF of X and Z :

$$f_{X,Z}(x, z) = f_{Z|X}(z | x) f_X(x) = f_Y(z - x) f_X(x). \tag{8.7}$$

To get the marginal f_Z , integrate over all possible x :

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Z}(x, z) dx = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx. \tag{8.8}$$

□

Why does the conditioning drop? I was unsure why we can remove “ $| X = x$ ”. Once $X = x$ is fixed, the only random variable left in $\{Y \leq z - x\}$ is Y , and Y is independent of X , so conditioning on $X = x$ does not change its distribution.

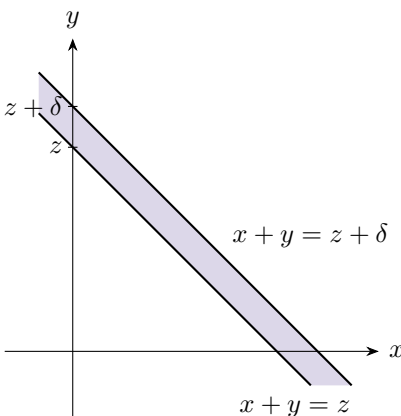
Proof by a thin strip. $\mathbb{P}(z \leq Z \leq z + \delta)$ is the probability that (X, Y) lands in the strip between the lines $y = z - x$ and $y = z + \delta - x$:

$$\begin{aligned}
 \mathbb{P}(z \leq Z \leq z + \delta) &= \int_{-\infty}^{\infty} \int_{z-x}^{z-x+\delta} f_{X,Y}(x, y) dy dx \\
 &= \int_{-\infty}^{\infty} \int_{z-x}^{z-x+\delta} f_X(x) f_Y(y) dy dx && \text{(independence)} \\
 &= \int_{-\infty}^{\infty} f_X(x) \left[\int_{z-x}^{z-x+\delta} f_Y(y) dy \right] dx \\
 &\approx \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) \delta dx,
 \end{aligned} \tag{8.9}$$

where the inner integral over an interval of width δ is approximated by a rectangle of height $f_Y(z - x)$. Since also $\mathbb{P}(z \leq Z \leq z + \delta) \approx f_Z(z) \delta$,

$$f_Z(z) \delta \approx \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) \delta dx, \tag{8.10}$$

and dividing by δ and letting $\delta \rightarrow 0$ gives (8.4). □



The strip has negative slope because $Z = X + Y$ means $Y = Z - X$. What we integrate over the strip is the joint *density*, so $\mathbb{P}(z \leq Z \leq z + \delta)$ is the probability mass of the strip, not its area.

8.3 Example: Sum of Two Uniforms

Example 8.2 (Sum of two independent uniforms)

Let $X, Y \sim \text{Uniform}[0, 1]$ be independent, so $f_X(x) = \frac{1}{b-a} = \frac{1}{1-0} = 1$ on $[0, 1]$, and the same for f_Y . Find f_Z for $Z = X + Y$.

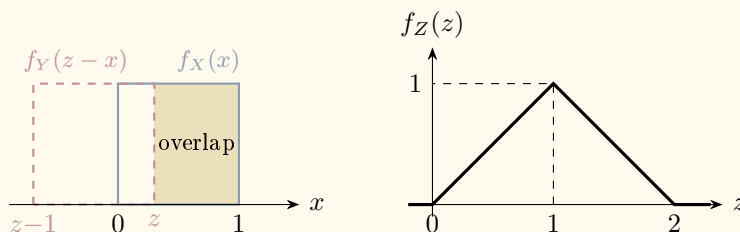
By (8.4), writing each density as an indicator,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{\infty} \mathbf{1}\{0 \leq x \leq 1\} \cdot \mathbf{1}\{0 \leq z-x \leq 1\} dx. \quad (8.11)$$

The second indicator is 1 when $z-1 \leq x \leq z$. So the integrand is 1 exactly on $[\max(0, z-1), \min(1, z)]$, and $f_Z(z)$ is the length of that interval:

$$\begin{aligned} 0 \leq z \leq 1: \quad f_Z(z) &= \int_0^z 1 dx = z, \\ 1 \leq z \leq 2: \quad f_Z(z) &= \int_{z-1}^1 1 dx = 1 - (z-1) = 2-z, \end{aligned} \quad (8.12)$$

and $f_Z(z) = 0$ otherwise. Check: the triangle has area $\frac{1}{2} \cdot 2 \cdot 1 = 1$.



Sliding overlap. As a function of x , $f_Y(z - x)$ is a box on $[z - 1, z]$ that slides to the right as z increases. $f_Z(z)$ is the length of its overlap with the box $[0, 1]$: it grows from 0 at $z = 0$, is largest at $z = 1$ (exact overlap), and shrinks back to 0 at $z = 2$. The same flip-and-slide picture is how convolution is used in signal processing.

8.4 Bounding Tail Probabilities

Example 8.3 (How big must a buffer be?)

Over n time slots, P_i packets arrive in slot i , where $P_1, \dots, P_n \sim \text{Poisson}(1)$ are i.i.d. A buffer holds $5n$ packets. How likely is it to overflow? We want to bound

$$\mathbb{P}\left(\sum_{i=1}^n P_i > 5n\right). \quad (8.13)$$

Theorem 8.2 (Markov's inequality)

If $X \geq 0$ and $a > 0$, then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \quad (8.14)$$

Proof. Pointwise, $X \geq a \mathbf{1}\{X \geq a\}$: if $X \geq a$ the right side is $a \leq X$, and otherwise it is $0 \leq X$. Taking expectations, $\mathbb{E}[X] \geq a \mathbb{E}[\mathbf{1}\{X \geq a\}] = a \mathbb{P}(X \geq a)$. $\square$

Let $S_n = \sum_{i=1}^n P_i \geq 0$. By linearity of expectation and $\mathbb{E}[P_i] = 1$ for $\text{Poisson}(1)$,

$$\mathbb{P}(S_n > 5n) \leq \mathbb{P}(S_n \geq 5n) \leq \frac{\mathbb{E}[S_n]}{5n} = \frac{\sum_{i=1}^n \mathbb{E}[P_i]}{5n} = \frac{\sum_{i=1}^n 1}{5n} = \frac{n}{5n} = \frac{1}{5}. \quad (8.15)$$

Too loose. The bound $\frac{1}{5}$ does not shrink as n grows, even though S_n/n stays close to its mean 1, far below 5, for large n . Markov only uses $\mathbb{E}[S_n]$.

The exponential trick. For $t > 0$, the function $x \mapsto e^{tx}$ is strictly increasing, so the event is unchanged by applying it to both sides. Apply Markov to the nonnegative random variable e^{tS_n} :

$$\mathbb{P}(S_n > 5n) = \mathbb{P}(e^{tS_n} > e^{5nt}) \leq \frac{\mathbb{E}[e^{tS_n}]}{e^{5nt}}. \quad (8.16)$$

Why e^{tx} ? I was confused why this particular function. We want to apply a function T so that $\mathbb{E}[T(\sum_i P_i)]$ splits into a product $\prod_i \mathbb{E}[T(P_i)]$ by independence. That requires $T(x+y) = T(x)T(y)$, which the exponential satisfies: $T(x+y) = e^{t(x+y)} = e^{tx}e^{ty} = T(x)T(y)$. The parameter $t > 0$ is a scaling factor we are free to choose.

Since the P_i are independent and identically distributed,

$$\mathbb{E}[e^{tS_n}] = \mathbb{E}[e^{tP_1+tP_2+\dots+tP_n}] = \mathbb{E}[e^{tP_1}] \mathbb{E}[e^{tP_2}] \dots \mathbb{E}[e^{tP_n}] = \left(\mathbb{E}[e^{tP_1}]\right)^n. \quad (8.17)$$

For $P \sim \text{Poisson}(\lambda)$, using the Taylor series $e^x = \sum_{k \geq 0} \frac{x^k}{k!}$,

$$\mathbb{E}[e^{tP}] = \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t \lambda)^k}{k!} = e^{-\lambda} e^{e^t \lambda} = e^{\lambda(e^t - 1)}. \quad (8.18)$$

Substituting into (8.16),

$$\mathbb{P}(S_n > 5n) \leq \frac{(e^{\lambda(e^t - 1)})^n}{e^{5nt}} = e^{\lambda n(e^t - 1) - 5nt} \stackrel{\lambda \equiv 1}{=} e^{n(e^t - 1 - 5t)}. \quad (8.19)$$

Choosing t . The bound holds for every $t > 0$, so minimize the exponent $g(t) = e^t - 1 - 5t$:

$$g'(t) = e^t - 5 = 0 \implies t^* = \ln 5, \quad g''(t) = e^t > 0, \quad (8.20)$$

$$g(t^*) = e^{\ln 5} - 1 - 5 \ln 5 = 4 - 5 \ln 5 \approx -4.05. \quad (8.21)$$

Therefore

$$\mathbb{P}\left(\sum_{i=1}^n P_i > 5n\right) \leq e^{-(5 \ln 5 - 4)n} \approx e^{-4.05n}. \quad (8.22)$$

For $n = 10$: Markov gives 0.2, while this bound gives $e^{-40.5} \approx 2.7 \times 10^{-18}$.

Is the bound $\propto e^{-n}$? I was unsure about the rate. Any $t > 0$ with $g(t) < 0$ gives exponential decay $e^{ng(t)}$; the best rate is $5 \ln 5 - 4 \approx 4.05$, at $t^* = \ln 5$. This method (Markov applied to e^{tX} , then optimizing over t) is called the **Chernoff bound**. The quantity $\mathbb{E}[e^{tP}]$ is the MGF of P , defined next.

8.5 Moment Generating Functions

Definition 8.2 (Moment generating function)

The **moment generating function** of a random variable X is

$$M_X(s) \triangleq \mathbb{E}[e^{sX}] = \begin{cases} \sum e^{sx} p_X(x) & (\text{discrete}), \\ \int_{-\infty}^{\infty} e^{sx} f_X(x) dx & (\text{continuous}), \end{cases} \quad (8.23)$$

for $s \in \mathbb{R}$ (it may equal ∞ for some s).

MGFs are useful for:

1. tight tail bounds (Section 8.4);
2. computing all the moments $\mathbb{E}[X^n]$ of a distribution;
3. sums of independent random variables: convolution becomes multiplication.

Theorem 8.3 (MGFs generate moments)

If $M_X(s) < \infty$ for all s in an open interval around 0, then for every $n \geq 1$,

$$M_X^{(n)}(0) = \mathbb{E}[X^n]. \quad (8.24)$$

Proof. Here s is a constant inside the expectation and X is the random part, so $\frac{d}{ds}e^{sX} = Xe^{sX}$. Differentiating n times and moving the derivative inside the expectation (allowed because M_X is finite near 0),

$$\frac{d^n}{ds^n}M_X(s) = \frac{d^n}{ds^n}\mathbb{E}[e^{sX}] = \mathbb{E}\left[\frac{d^n}{ds^n}e^{sX}\right] = \mathbb{E}[X^n e^{sX}]. \quad (8.25)$$

At $s = 0$, $e^{0 \cdot X} = e^0 = 1$, so $M_X^{(n)}(0) = \mathbb{E}[X^n e^{sX}]|_{s=0} = \mathbb{E}[X^n]$. □

Example 8.4 (MGF of the exponential)

Let $X \sim \text{Exponential}(\lambda)$: $f_X(x) = \lambda e^{-\lambda x}$ for $x \in [0, \infty)$. Then

$$M_X(s) = \int_0^\infty e^{sx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{(s-\lambda)x} dx. \quad (8.26)$$

If $s < \lambda$, the exponent $s - \lambda$ is negative and

$$M_X(s) = \lambda \left[\frac{1}{s-\lambda} e^{(s-\lambda)x} \right]_0^\infty = \lambda \left(0 - \frac{1}{s-\lambda} \right) = \frac{\lambda}{\lambda-s}. \quad (8.27)$$

If $s \geq \lambda$, the integrand $e^{(s-\lambda)x} \geq 1$ for all $x \geq 0$, so the integral diverges:

$$M_X(s) = \begin{cases} \frac{\lambda}{\lambda-s} & s < \lambda, \\ \infty & s \geq \lambda. \end{cases} \quad (8.28)$$

Moments. $M'_X(s) = \frac{\lambda}{(\lambda-s)^2}$ and $M''_X(s) = \frac{2\lambda}{(\lambda-s)^3}$, so

$$\mathbb{E}[X] = M'_X(0) = \frac{1}{\lambda}, \quad \mathbb{E}[X^2] = M''_X(0) = \frac{2}{\lambda^2}, \quad \text{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \quad (8.29)$$

Theorem 8.4 (MGFs determine the distribution)

If $M_X(s) = M_Y(s) < \infty$ for all s in some open interval $(-\varepsilon, \varepsilon)$ around 0, then X and Y have the same distribution.

So if we recognize an MGF, we know the distribution. (The proof is beyond the scope of this course.)

Not finished in lecture. We ran out of time after starting the list of MGF properties. The main properties are filled in below.

Theorem 8.5 (Properties of MGFs)

1. $M_X(0) = 1$.
2. If $Y = aX + b$, then $M_Y(s) = e^{sb} M_X(as)$.
3. If X and Y are independent, then $M_{X+Y}(s) = M_X(s) M_Y(s)$. More generally, for independent $X_1, \dots, X_n$, $M_{X_1+\dots+X_n}(s) = \prod_{i=1}^n M_{X_i}(s)$, and $(M_{X_1}(s))^n$ if they are also identically distributed.

Proof. 1. $M_X(0) = \mathbb{E}[e^{0 \cdot X}] = \mathbb{E}[1] = 1$.

2. $M_Y(s) = \mathbb{E}[e^{s(aX+b)}] = \mathbb{E}[e^{sb} e^{(as)X}] = e^{sb} \mathbb{E}[e^{(as)X}] = e^{sb} M_X(as)$, since e^{sb} is a constant.
3. $M_{X+Y}(s) = \mathbb{E}[e^{s(X+Y)}] = \mathbb{E}[e^{sX} e^{sY}] = \mathbb{E}[e^{sX}] \mathbb{E}[e^{sY}] = M_X(s) M_Y(s)$, using independence of e^{sX} and e^{sY} . The n -term version follows by induction. $\square$

Property 3 is the payoff: the density of $X + Y$ requires a convolution integral, but its MGF is just a product.

8.6 MGF Proofs for Sums of Independent Random Variables

Lemma 8.1 (MGF of the Gaussian)

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then for all $s \in \mathbb{R}$,

$$M_X(s) = e^{\mu s + \sigma^2 s^2 / 2}. \quad (8.30)$$

Proof. First let $Z \sim \mathcal{N}(0, 1)$. Complete the square in the exponent, $z^2 - 2sz = (z - s)^2 - s^2$:

$$\begin{aligned} M_Z(s) &= \int_{-\infty}^{\infty} e^{sz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z^2 - 2sz)/2} dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-s)^2/2 + s^2/2} dz \\ &= e^{s^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-s)^2/2} dz = e^{s^2/2} \cdot 1, \end{aligned} \quad (8.31)$$

since the last integrand is the $\mathcal{N}(s, 1)$ density. Now $X = \sigma Z + \mu$ (Lecture 7), so by Property 2 with $a = \sigma$, $b = \mu$,

$$M_X(s) = e^{\mu s} M_Z(\sigma s) = e^{\mu s} e^{\sigma^2 s^2 / 2} = e^{\mu s + \sigma^2 s^2 / 2}. \quad (8.32)$$

$\square$

Theorem 8.6 (Sum of independent Gaussians)

If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ are independent, then

$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2). \quad (8.33)$$

Proof. By Property 3 and the Gaussian MGF,

$$\begin{aligned} M_{X+Y}(s) &= M_X(s) M_Y(s) = e^{\mu_X s + \sigma_X^2 s^2 / 2} \cdot e^{\mu_Y s + \sigma_Y^2 s^2 / 2} \\ &= e^{(\mu_X + \mu_Y)s + (\sigma_X^2 + \sigma_Y^2)s^2 / 2}, \end{aligned} \quad (8.34)$$

which is the MGF of $\mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$, finite for all s . Since MGFs determine the distribution, $X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$. $\square$

This completes the proof left as TBD in Lecture 7.

Corollary 8.1 (Sum of independent Poissons)

If $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ are independent, then $X + Y \sim \text{Poisson}(\lambda + \mu)$.

Proof. By (8.18), $M_X(s) = e^{\lambda(e^s-1)}$ and $M_Y(s) = e^{\mu(e^s-1)}$, so

$$M_{X+Y}(s) = e^{\lambda(e^s-1)} e^{\mu(e^s-1)} = e^{(\lambda+\mu)(e^s-1)}, \quad (8.35)$$

the MGF of $\text{Poisson}(\lambda + \mu)$. (Compare the convolution proof of Poisson merging in Lecture 4.) $\square$

End of Lecture 8.