\[ E[g(x)] = \int g(x) f_x(x) \, dx. \]

\[ X \sim f_x(x). \]

\[ Y = g(x). \]

\[ Y \sim \frac{f_x(x)}{|\frac{dy}{dx}|}. \]

**Proof:**

Find CDF of \( Y \), then take derivatives

\[ P(Y \leq y) = P(g(x) \leq y) \]

\[ = P(X \leq g^{-1}(y)) \]

\[ = F_x(g^{-1}(y)) = F_x(x). \]

Take derivatives with \( y \).

\[ f_Y(y) = f_X(x) \frac{dx}{dy}. \]

\[ \int f_x(x) \, dx = f_Y(y) \, dy. \]
\[ T = X + Y \]
\[ E[T] = E[X] + E[Y] \]

\[ f_T(t) = ? \]
Assume \( X, Y \) independent.

**Discrete Case**

\[ P(T = k) = \sum_{x=0}^{k} P(X = x) P(Y = k - x) \]

**Continuous**

\[ f_T(t) = \int_{-\infty}^{\infty} f_X(x) f_Y(t-x) \, dx \]

**Convolution**
\( CDF \quad F_T(t) = P(T \leq t) \)

\[
= \int_{-\infty}^{\infty} P(x + y \leq t \mid x = x)f_x(x) \, dx
\]

\[
= \int_{-\infty}^{\infty} P(y \leq t - x)f_x(x) \, dx
\]

Let P.

Substitute \( x = \infty \) and drop the conditioning as \( x, y \) are independent

\[
F_T(t) = \int_{-\infty}^{\infty} F_y(t - x)f_x(x) \, dx.
\]

Take derivatives w.r.t. \( t \)

\[
F_T(t) = \int_{-\infty}^{\infty} f_x(x)f_y(t - x) \, dx.
\]

Example \( X \sim N(0, 1) \)
\( Y \sim N(0, 1) \) independent.

\( T = X + Y \)

\[
F_T(t) = \int_{-\infty}^{\infty} f_x(x)f_y(t - x) \, dx
\]
\[
\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{x^2}{2} \right) \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{(x-\xi)^2}{2} \right)
\]

\[
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left( -\frac{1}{2} \left[ x^2 + \xi^2 - 2\xi x + x^2 \right] \right)
\]

\[
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left( -\frac{1}{2} \left[ 2x^2 - 2\xi x + \xi^2 \right] \right) dx
\]

\[
2x^2 - 2\xi x + \xi^2
\]

\[
2 \left( x^2 - \xi x + \frac{\xi^2}{2} \right)
\]

\[
2 \left[ (x - \frac{\xi}{2})^2 - \frac{\xi^2}{4} + \frac{\xi^2}{2} \right]
\]

\[
\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} \left[ (x - \frac{\xi}{2})^2 + \frac{\xi^2}{4} \right] \right) dx
\]

\[
= \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{\xi^2}{4} \right) \int_{-\infty}^{\infty} \exp\left( -\frac{1}{2} \left( x - \frac{\xi}{2} \right)^2 \right) dx
\]

This looks like the
\[
\int_{-\infty}^{\infty} \exp \left( -\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} \right) \, dx = \sqrt{2\pi} \sigma
\]

\[
\mu = \frac{\epsilon}{2}
\]
\[
\sigma^2 = \frac{1}{2}
\]

So,
\[
\int_{-\infty}^{\infty} \exp \left( -\frac{1}{2} \frac{\epsilon^2}{x^2} \right) \, dx = \frac{1}{\sqrt{2}} \sqrt{2\pi} = \sqrt{\pi}
\]

\[
\phi_T(t) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{\epsilon^2}{4} \right) \cdot \sqrt{\frac{2\pi}{\sqrt{2}}}
\]

\[
= \frac{1}{\sqrt{2} \sqrt{2\pi}} \exp \left( -\frac{1}{2} \frac{\epsilon^2}{2} \right) = N(0, 2)
\]

Hence, the sum of two independent \( N(0, 1) \) RVs is an \( N(0, 2) \) RV.

In general, any linear combination of Normal RVs will be Normal.

\[
\begin{align*}
\begin{bmatrix} x \sim N(\mu_1, \sigma_1^2) \\
y \sim N(\mu_2, \sigma_2^2) \end{bmatrix} & \Rightarrow x + y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)
\end{align*}
\]
Poisson:

\( X \sim \text{Pois} (\lambda) \)

Mean = \( \lambda \)

Variance = \( \lambda \)

\[
E[X^2] = \sum_{k=0}^{\infty} k^2 p(X = k)
\]

\[
= \sum_{k=0}^{\infty} k^2 \frac{e^{-\lambda} \lambda^k}{k!}
\]

\[
= \lambda^2 + \lambda
\]

Binomial:

\( X \sim \text{Bin} (n, p) \)

\[
E[X^2] = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k}
\]

\( X \) is sum of \( n \) iid Bernoulli \( (p) \) RVs

Var of a sum of independent RVs = sum of variances of the RVs. (Proof later)
\[ X = X_1 + X_2 + X_3 + \ldots + X_n \]

\[ = I_1 + I_2 + I_3 + \ldots + I_n \]

\[ X^2 = (I_1 + I_2 + \ldots + I_n)(I_1 + I_2 + \ldots + I_n) \]

\[ = I_1^2 + I_2^2 + \ldots + I_n^2 \]

\[ + 2 I_1 I_2 + 2 I_1 I_3 + \ldots + 2 I_{n-1} I_n \]

\[ E[X^2] = n E[I_i^2] + 2 \binom{n}{2} E[I_i I_j] \]

\[ \text{each } I_i^2 \text{ has the same distribution} \]

\[ I_i - \text{indicator variable of success on 1st trial}. \]

\[ I_i^2 - \text{takes the values 0, 1 with the same probability that } I_i \text{ does.} \]

Expected value of an indicator RV

- prob. of the event that it is an indicator for.

\[ E[I] = 0 \times p(I = 0) + 1 \times p(I = 1) \]

\[ \Rightarrow E[I^2] = p \]
\[ E[I_{1}I_{2}] = p^{2} \]

\[ E[x^{2}] = np + n(n-1)p^{2} \]
\[ = np + np^{2} - np^{2} \]
\[ = np - np^{2} \]
\[ = np(1-p) \]
\[ = npq. \]
What can we predict about the store based on what we have observed so far?

What is the probability that the sun will rise tomorrow?

The sun has risen for the last n days.

 labeled: \( X_1, X_2, \ldots, X_n \) iid Bernoulli (\( \theta \)).

What I'm interested in is the unknown success probability \( \theta \).

\( \theta \) is a random variable.

It has a distribution \( f_\theta(x) \).

\( f_\theta(x) \) encodes our knowledge of the success prob. of the Bernoulli distribution.
\[ f_{\theta \mid d} (\theta \mid D = d) \]

\[ = \frac{f_{d \mid \theta} (d \mid \theta) f_\theta (\theta)}{f_d (d)} \]

Continuous form of Bayes theorem:

\[ f_d (d) = \int f_{d \mid \theta} (d \mid \theta) f_\theta (\theta) \, d\theta \]

Thus it is not a function of \( \theta \).

- Sometimes we ignore it, if we're interested in the shape of \( f_{\theta \mid d} \).
- Sometimes we can recognize the numerator as a distribution we know about, and so we know directly how it is normalized.
Before we collect data.

$U(0,1)$ — we have no idea what values of $\Theta$ (success probability) we likely before collecting data.

\[ \mathbb{E} \int \theta \cdot d\theta \]

Assume a value for the success prob. $\Theta$.

What's the probability of observing the data?

$n$ successes in $n$ trials

\[ p(\text{data} | \theta) = \theta^n \]

\[ p(\theta | \text{data}) \propto \theta^n = c \theta^n \]

Hence

\[ \int_0^1 c \theta^n \, d\theta = 1. \]

\[ \left. c \frac{\theta^{n+1}}{n+1} \right|_0^1 = 1 \]

\[ c = n+1 \]
This summarises our knowledge of
the success prob. after having observed
the data.

What's the prob. that the sun will rise
tomorrow?

- prob. of success on \((n+1)\text{th}\) trial, given
data for \(1\text{st} \ldots n\) trials.

\[
\begin{align*}
  & P(X_{n+1} = 1 \mid X_1, X_2 \ldots X_n) \\
  &= \int_\Theta P(X_{n+1} = 1 \mid \Theta) \cdot f_{\Theta|D}(\Theta \mid d) \, d\Theta \quad \text{(LOTP)} \\
  &= \int_0^1 \Theta (n+1) \Theta^n \, d\Theta. \\
  &= \frac{n+1}{n+2}. \quad \text{Predictive probability.}
\end{align*}
\]

(prob. of some unobserved
events based on the event
... have observed)