PLSC 502 – Fall 2023
Probability and Random Variables

September 25, 2023

Terminology

Some concepts:

- **X**: A *random variable*
- **Outcome**: A possible event / result of a process
- **Realization**: One observation of the process (x)
- **Sample Space (S)**: The set of all possible outcomes

What is a "Random Variable"?

Formally:

$$X : S \to M$$

where M is a *measurable space*. Often:

$$X : S \to \mathbb{R}$$

where $\mathbb{R}$ are the real numbers.

More informally, a random variable is a formalization of an event or process in which the result is subject to unmeasured (and unaccounted-for) variation.

Sample Spaces

For a *discrete* variable X:

$$X \in S = \{x_1, x_2, ...x_J\}$$

For a *continuous* X:

$$X \in S = [\ell, \tau].$$

E.g., for points in the English Premier League (match week five, that we discussed last class):

$$X \in S = \{0, 1, 2, ...14, 15\}$$

and a realization of that random variable is:

$$X_{\text{Spurs}} = 13.$$

Probability (*Frequentist*)

Probability $=$ *Long-run relative frequency*.

$$\Pr(\text{Event}) = \frac{\text{The number of times the } \textit{event of interest} \text{ can or could occur}}{\text{The number of times } \textit{any event} \text{ can or could occur}}.$$

More formally:

$$\Pr(X = x) = \lim_{N \to \infty} \left(\frac{\sum_N I\{X_i = x\}}{N} \right)$$

where $I\{\cdot\}$ is an *indicator function* for $X_i = x$.

Probability: Characteristics

For anything that is a probability:

- It's value necessarily **ranges between zero and one**:

$$\Pr(X = x) \in [0, 1].$$

- The **sum of probabilities for all outcomes always equals one**:

$$\sum_{j=1}^{J} \Pr(X = x_j) \equiv \Pr(S) = 1.0$$

The Multiplication Rule

The probability of obtaining a *combination* of independent, mutually exclusive outcomes is equal to the *product* of their separate probabilities.

Formally:

$$\Pr(X = x_j \cap X = x_\ell) = \Pr(X = x_j) \times \Pr(X = x_\ell),\ j \neq \ell$$

The Addition Rule

The probability of obtaining *any one* (or more) of several independent, mutually exclusive outcomes is equal to the *sum* of the probabilities for those events.

Formally:

$$\Pr(X = x_j \cup X = x_\ell) = \Pr(X = x_j) + \Pr(X = x_\ell),\ j \neq \ell$$

Addition Rule (continued)

If events are not mutually exclusive:

$$\begin{aligned}\Pr(X = x_j \cup X = x_\ell) &= \Pr(X = x_j) + \Pr(X = x_\ell) \\ &\quad - \Pr(X = x_j \cap X = x_\ell)\end{aligned}$$

So, for example, Pr(Diamond *or* face-card):

$$\begin{aligned}\Pr(Z) &= \Pr(\text{Diamond}) + \Pr(\text{Face-Card}) \\ &\quad - \Pr(\text{Diamond-Suited Face Card}) \\ &= \frac{1}{4} + \frac{12}{52} - \frac{3}{52} \\ &= 0.25 + 0.23 - 0.06 \\ &= \mathbf{0.42}\end{aligned}$$

Independence

Consider $\Pr(X = x_j, X = x_\ell) = \Pr(X = x_j \ \cap \ X = x_\ell)$ ("joint PDF")...

If x_j and x_ℓ are *independent*:

- $\Pr(X = x_j, X = x_\ell) = \Pr(X = x_j) \times \Pr(X = x_\ell)$.
- That is, *the joint PDF is equal to the product of the marginal PDFs.*
- We write $X_j \perp X_\ell$.

Conditional Probability

If x_j and x_ℓ are *not independent*...

Conditional probabilities:

- I.e., $\Pr(X = x_j | X = x_\ell)$ and/or $\Pr(X = x_\ell | X = x_j)$
- Say "The probability of x_j given x_ℓ," etc.

Implies:

$$\begin{aligned}\Pr(X = x_j | X = x_\ell) &= \frac{\Pr(X = x_j, X = x_\ell)}{\Pr(X = x_\ell)}, \text{ and} \\ \Pr(X = x_\ell | X = x_j) &= \frac{\Pr(X = x_j, X = x_\ell)}{\Pr(X = x_j)}\end{aligned}$$

Independence, Defined

If two variables are *independent*, then:

$$
\begin{aligned}
\Pr(X = x_j | X = x_\ell) &= \frac{\Pr(X = x_j, X = x_\ell)}{\Pr(X = x_\ell)} \\
&= \frac{\Pr(X = x_j) \times \Pr(X = x_\ell)}{\Pr(X = x_\ell)} \\
&= \Pr(X = x_j)
\end{aligned}
$$

Holds for any number of realizations; e.g. for x_j, x_ℓ, and x_k:

$$
\begin{aligned}
\Pr(X = x_j, X = x_\ell, X = x_k) &= \Pr(X = x_j | X = x_\ell, X = x_k) \\
&\quad \times \Pr(X = x_\ell | X = x_k) \\
&\quad \times \Pr(X = x_k)
\end{aligned}
$$

Because $\Pr(X = x_j, X = x_\ell) = \Pr(X = x_\ell, X = x_j)$, we can write:

$$\begin{aligned}\Pr(X = x_j | X = x_\ell) \times \Pr(X = x_\ell) &= \Pr(X = x_\ell | X = x_j) \\ &\quad \times \Pr(X = x_j)\end{aligned}$$

and so:

$$\Pr(X = x_j | X = x_\ell) = \frac{\Pr(X = x_\ell | X = x_j) \times \Pr(X = x_j)}{\Pr(X = x_\ell)}.$$

More Bayes' Rule

Generally:

$$\Pr(A|B) = \frac{\Pr(B|A) \times \Pr(A)}{\Pr(B)}$$

Informally:

$$\text{Posterior} = \frac{\text{Likelihood} \ \times \ \text{Prior}}{\text{Marginal}}$$

Probability: Frequentist and Bayesian

"Frequentist": Probability is the long-run relative frequency of an event.

- Sometimes: "physical" probability (a la a physical system) or "objective"
- E.g., Laplace, Neyman, Pearson, Fisher

"Bayesian": Probability is the best subjective belief about the state of an event.

- Sometimes: "epistemic" (or "subjective") probability
- E.g., Savage, de Finetti, Jeffreys, Wald

Probability and Odds

Odds are a ratio:

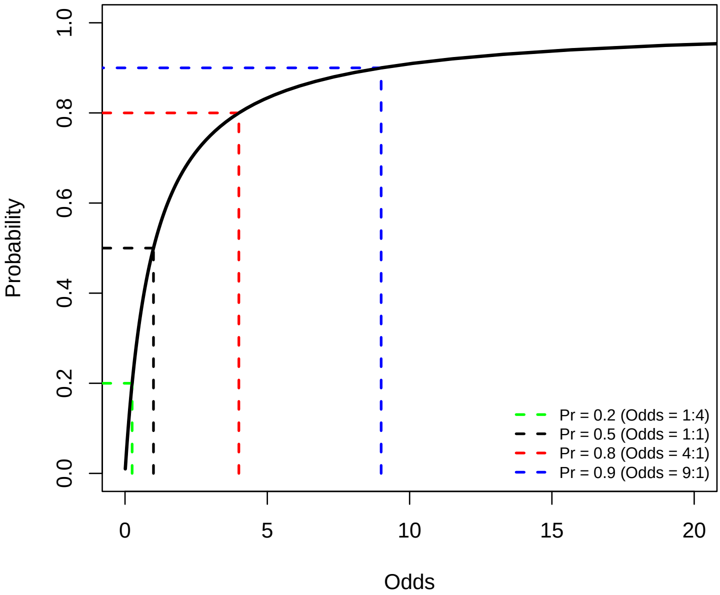
$$\begin{aligned}\text{Odds}(X = x_j) &= \frac{\Pr(X = x_j)}{\Pr(X \neq x_j)} \\ &= \frac{\Pr(X = x_j)}{1 - \Pr(X = x_j)}\end{aligned}$$

Often written as $\Pr(X = x_j) : [1 - \Pr(X = x_j)]$.

E.g., "The odds of x_j are 4:1 (in favor)":

- $\Pr(X = x_j) = \frac{4}{4+1} = 0.8$
- $\Pr(X \neq x_j) = \frac{1}{4+1} = 0.2$

Probability and Odds



Log-Odds

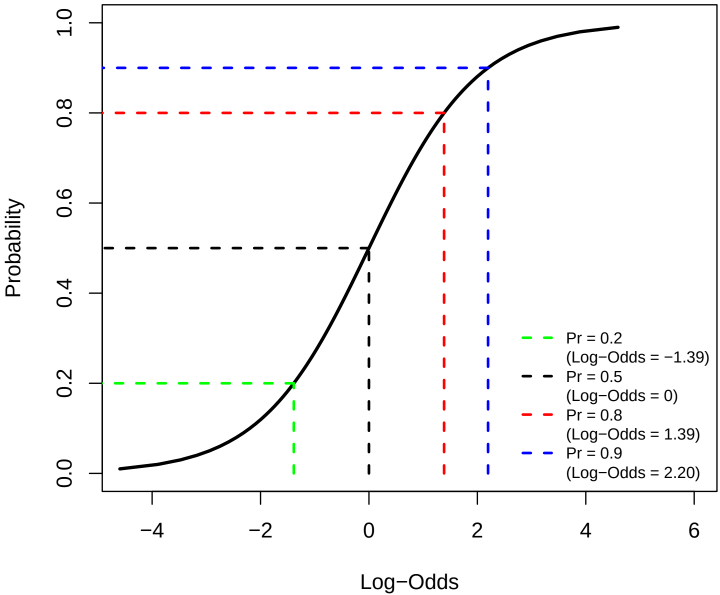
Log-odds:

$$
\begin{aligned}
\ln[\text{Odds}(X = x_j)] &= \ln\left[\frac{\Pr(X = x_j)}{\Pr(X \neq x_j)}\right] \\
&= \ln\left[\frac{\Pr(X = x_j)}{1 - \Pr(X = x_j)}\right]
\end{aligned}
$$

Note that:

- Odds $\in\ [0, \infty)$, but
- Log-odds $\in\ (-\infty, \infty)$.

Log-Odds and Probability



Likelihood

For N realizations of X:

$$\begin{aligned} X_1 &= x_1 \\ X_2 &= x_2 \\ X_3 &= x_3 \\ &\vdots \\ X_N &= x_N \end{aligned}$$

Likelihood:

$$L(X) = \Pr(X_1 = x_1, X_2 = x_2, \ldots X_N = x_N)$$

Likelihood (continued)

If $X_j \perp X_k \; \forall \; j, k$ then:

$$\begin{aligned} L(X) &= \Pr(X_1 = x_1) \times \Pr(X_2 = x_2) \times \ldots \times \Pr(X_N = x_N) \\ &= \prod_{i=1}^{N} \Pr(X_i = x_i). \end{aligned}$$

Log-Likelihood:

$$\begin{aligned} \ln L(X) &= \ln \left[\prod_{i=1}^{N} \Pr(X_i = x_i) \right] \\ &= \sum_{i=1}^{N} \ln[\Pr(X_i = x_i)] \end{aligned}$$

Random Variables

Continuous and Discrete Variables

Discrete Variables

- $X \in S = \{s_0, s_1, ...\}$
- $\Pr(s) \geq 0$ for each $s \in S$
- $\sum_{s \in S} \Pr(s) = 1$

Continuous Variables

- $X \in S \in \Re$
- $\exists\ f(x)$ such that for any closed interval $[a, b]$ $\Pr(a < x \leq b) = \int_b^a f(x)dx$.
- Requires:
  - $f(x) \geq 0$ for all x
  - $\int_{-\infty}^{\infty} f(x)dx = 1$

Probability Density Function

The PDF is the function $f(x)$ that maps the possible values of X to some associated probability of their occurrence.

Discrete X:

$$f(x) = \Pr(X = x) \, \forall \, x \in S$$

Continuous X:

$$\Pr(a < X \leq b) = \int_a^b f(x)\, dx$$

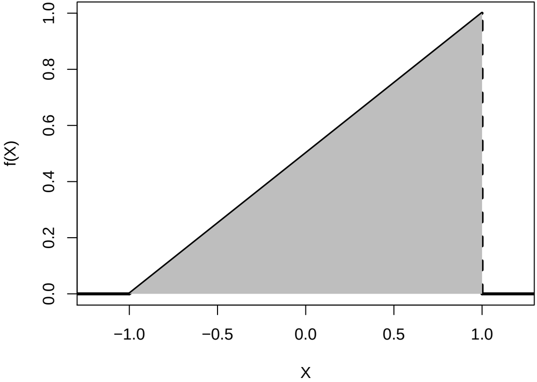
Again: Requires:

- $f(x) \geq 0$ for all x
- $\int_{-\infty}^{\infty} f(x) dx = 1$

An Example

Consider:

$$f(X) = \begin{cases} \frac{X+1}{2} & \text{for}[-1 \leq x < 1], \\ 0 & \text{otherwise.} \end{cases}$$



Is It A PDF?

1. Is $f(X) \geq 0 \; \forall \; x$? – Yes.
2. Is

$$\begin{aligned}\Pr(-\infty \leq x \leq \infty) &\equiv \Pr(-1 \leq x \leq 1) \\ &= \int_{-1}^{1} f(x)dx = 1?\end{aligned}$$

Let's see:

$$\begin{aligned}\Pr(-1 \leq x \leq 1) &= \int_{-1}^{1} \frac{1}{2}(x+1)dx \\ &= \frac{1}{2}\left(\frac{X^2}{2} + x\right)\Bigg|_{-1}^{1} \\ &= \frac{1}{2}\left(\frac{1^2}{2} + 1\right) - \frac{1}{2}\left(\frac{-1^2}{2} - 1\right) \\ &= 0.75 - (-0.25) \\ &= 1\end{aligned}$$

Cumulative Distribution Function (CDF)

The CDF is the probability that X will take on a value less than or equal to than some value x in its range.

Discrete X:

$$\begin{aligned} \Pr(X \leq x) \equiv F(x) &= \sum_{X \leq x} \Pr(X = x) \\ &= 1 - \sum_{X > x} \Pr(X = x) \end{aligned}$$

CDF (continued)

Continuous X:

$$\Pr(X \leq x) \equiv F(x) = \int_{-\infty}^{x} f(t)\, dt$$

Properties:

- $0 \leq F(x) \leq 1$.
- Nondecreasing in X.
- $\Pr(x > k) = 1$ - F(k).
- $\Pr(a < x \leq b) = F(b) - F(a)$.
- $F(-\infty) = 0$.
- $F(\infty) = 1$.

Example, Again

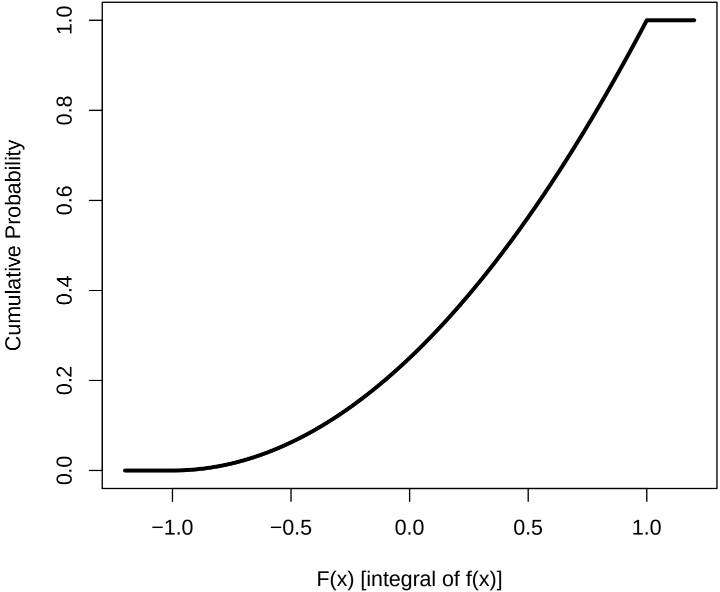
For:

$$f(X) = \begin{cases} \frac{1}{2}(x+1) & \text{for} -1 \leq x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

we already know that $\int_{-1}^{1} f(x)dx = 1$.

$$\begin{aligned} F(x) &= \int_{-1}^{1} f(t)dt \\ &= \int_{-1}^{1} \frac{1}{2}(t+1)dt \\ &= \frac{1}{2}\left(\frac{t^2}{2}+t\right)\Bigg|_{-1}^{1} \\ &= \frac{1}{2}\left(\frac{t^2}{2}+t\right)+c. \end{aligned}$$

Example CDF, Illustrated



Expected Value

For X with PDF $f(x)$ and CDF $F(x) = \int_{-\infty}^{x} f(t)\,dt$, the *expected value* of X [denoted $\mathrm{E}(X)$, or μ] is the *probability-weighted mean of the potential values of that variable*.

<u>Discrete X</u>:

$$\mathrm{E}(X) = \sum_{x} [x \times f(x)]$$

E.g., number of heads in two coin flips:

0 Heads	Prob. $= .25$	Prob$\times$Value $= .25\times 0$	$=$	0
1 Head	Prob. $= .50$	Prob$\times$Value $= .50\times 1$	$=$	.50
2 Heads	Prob. $= .25$	Prob$\times$Value $= .25\times 2$	$=$	.50
		$\sum$	$=$	1.0

Expected Value (continued)

Continuous X:

$$E(X) = \int [x \times f(x)] dx$$

Properties:

- $E(c) = c$
- $E(x + y + z) = E(x) + E(y) + E(z)$
- If $g(x)$ is some function of x, then

$$\begin{aligned} E[g(x)] &= \sum [g(x) \times \text{Prob}(X = x)] \, \forall x \text{ (discrete case)} \\ &= \int g(x) f(x) \, dx \text{ (continuous case)} \end{aligned}$$

- This includes a constant function: $E(cx) = cE(x)$.
- Implies that for $g(x) = a + bx$, $E(a + bx) = a + bE(x)$.

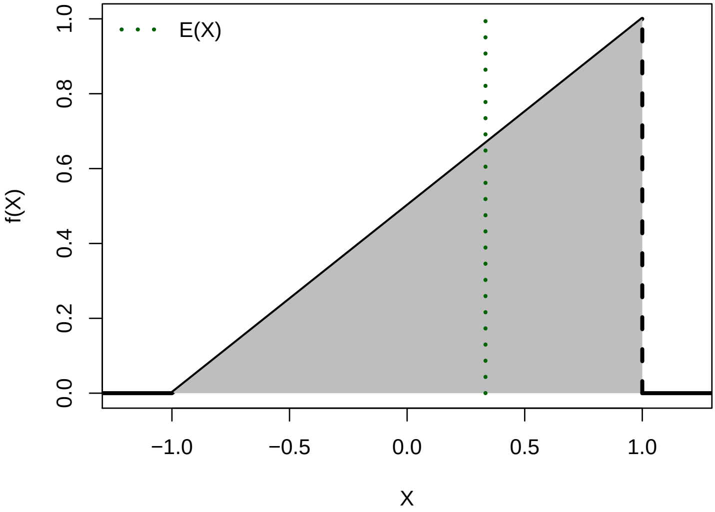
Example Again

For random variable X with

$$f(x) = \begin{cases} \frac{1}{2}(x+1) & \text{for} -1 \leq y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

What is $E(X)$?

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-1}^{1} x \left(\frac{x+1}{2} \right) dx \\
&= \int_{-1}^{1} \frac{1}{2}(x^2 + x)\, dx \\
&= \frac{1}{2} \int_{-1}^{1} x^2 \, dx + \frac{1}{2} \int_{-1}^{1} x \, dx \\
&= \frac{1}{2} \left(\frac{x^3}{3} + c_1 \right) \Bigg|_{-1}^{1} + \frac{1}{2} \left(\frac{x^2}{2} + c_2 \right) \Bigg|_{-1}^{1} \\
&= \frac{1}{2} \left(\frac{x^3}{3} + \frac{x^2}{2} + c_3 \right) \Bigg|_{-1}^{1} \\
&= \frac{1}{2} \left[\left(\frac{1^3}{3} + \frac{1^2}{2} + c_3 \right) - \left(\frac{-1^3}{3} + \frac{-1^2}{2} + c_3 \right) \right] \\
&= \frac{1}{3}
\end{aligned}$$


E(X)
1.0
0.8
0.6
0.4
0.2
0.0
f(X)
−1.0
−0.5
0.0
0.5
1.0
X

Variance

Generally:

$$\mathrm{Var}(X) = \mathrm{E}[(x - \mu)^2]$$

Discrete X:

$$\mathrm{Var}(X) = \sum (x - \mu)^2 f(x)$$

Continuous X:

$$\mathrm{Var}(X) = \int (x - \mu)^2 f(x)\, dx$$

Variance (continued)

$$
\begin{aligned}
\mathrm{E}[(x-\mu)^2] \equiv \sigma^2 &= \mathrm{E}[x^2 - 2x\mu + \mu^2] \\
&= \mathrm{E}(x^2) - 2\mu\mathrm{E}(x) + \mathrm{E}(\mu^2) \\
&= \mathrm{E}(x^2) - 2\mu^2 + \mu^2 \\
&= \mathrm{E}(x^2) - \mu^2 \\
&\equiv \left(\int x^2 f(x)\, dx - \mu^2\right)
\end{aligned}
$$

- We often write the variance as σ^2, and the positive square root of it (the standard deviation) as σ.
- This also implies that the expectation of the square of a variable X is $\mathrm{E}(x^2) = \sigma^2 + \mu^2$.

Variance Properties

- $\text{Var}(X) > 0$, except
- $\text{Var}(c) = 0$
- $\text{Var}(cX) = c^2\text{Var}(X)$
- $\text{Var}(a + bX) = b^2\text{Var}(X)$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$

Example Again

What is the variance of $f(X) = \frac{1}{2}(x+1)$ for the range $-1 < x < 1$?

Recall that $\mu = 1/3$, so:

$$
\begin{aligned}
\mathrm{E}[(x-\mu)^2] \equiv \sigma^2 &= \int_{-1}^{1} X^2 f(x)\,dx - \mu^2 \\
&= \int_{-1}^{1} \frac{1}{2}x^2(x+1)\,dx - \left(\frac{1}{3}\right)^2 \\
&= \frac{1}{2}\left(\frac{x^4}{4} + c_1\right)\Big|_{-1}^{1} + \frac{1}{2}\left(\frac{x^3}{3} + c_2\right)\Big|_{-1}^{1} - \frac{1}{9} \\
&= \frac{1}{2}\left(\frac{X^4}{4} + \frac{X^3}{3} + c_3\right)\Big|_{-1}^{1} - \frac{1}{9} \\
&= \frac{1}{2}\left[\left(\frac{1^4}{4} + \frac{1^3}{3} + c_3\right) - \left(\frac{-1^4}{4} + \frac{-1^3}{3} + c_3\right)\right] - \frac{1}{9} \\
&= \frac{19}{72}\ (\approx 0.2639).
\end{aligned}
$$

Moments, redux

The kth moment of X is:

$$M_k = \mathrm{E}(X^k)$$

The kth moment exists if:

$$\begin{aligned} \mathrm{E}(|X|^k) &< \infty \\ = \int_{-\infty}^{\infty} |x|^k f(x) dx &< \infty \text{ (for continuous } X) \end{aligned}$$

"Central" moments:

$$\mu_k = \mathrm{E}[(X - \mu)^k]$$

Moment-Generating Functions

For $t \in \mathbb{R}$,

$$\psi(t) = \mathrm{E}[\exp(tX)]$$

For continuous X:

$$\begin{aligned}
\psi(t) &= \int_{-\infty}^{\infty} \exp(tx) f(x) dx \\
&= \int_{-\infty}^{\infty} \left(1 + tx + \frac{t^2 x^2}{2!} + \dots\right) f(x) dx \\
&= 1 + t\mathrm{E}(X) + \frac{t^2 \mathrm{E}(X^2)}{2} + \dots \\
&= 1 + tM_1 + \frac{t^2 M_2}{2} + \dots
\end{aligned}$$

Note that:

$$\begin{aligned}
\psi(0) &= \mathrm{E}[\exp(0)] \\
&= 1
\end{aligned}$$

MGFs Can Be Useful

First:

$$\psi(t) = \int_{-\infty}^{\infty} \exp(tx)\, d\, F(x).$$

Second:

$$\begin{aligned}
\left.\frac{\partial^k \psi(t)}{\partial^k t}\right|_{t=0} &= \left.\frac{\partial^k \mathsf{E}[\exp(tX)]}{\partial^k t}\right|_{t=0} \\
&= \mathsf{E}\left[\left.\frac{\partial^k \exp(tX)}{\partial^k t}\right|_{t=0}\right] \\
&= \mathsf{E}\{[X^k \exp(tX)]|_{t=0}\} \\
&= \mathsf{E}(X^k)
\end{aligned}$$

Next time: Probability Distributions