Lecture 3: The Distribution Zoo

PhD Mathematics II: Probability

Juan Llavador Peralt

IIES

November 6, 2024

Outline

Models for Discrete Random Variables

Models for Continuous Random Variables

Functions of Random Variables

The Distribution Zoo

V · T · E	Probability distributions (list)	[hide]
Discrete univariate	with finite support	Benford · Bernoulli · Beta-binomial · Binomial · Categorical · Hypergeometric (Negative) · Poisson binomial · Rademacher · Soliton · Discrete uniform · Zipf · Zipf–Mandelbrot
	with infinite support	Beta negative binomial · Borel · Conway–Maxwell–Poisson · Discrete phase-type · Delaporte · Extended negative binomial · Flory–Schulz · Gauss–Kuzmin · Geometric · Logarithmic · Mixed Poisson · Negative binomial · Panjer · Parabolic fractal · Poisson · Skellam · Yule–Simon · Zeta
Continuous univariate	supported on a bounded interval	Arcsine · ARGUS · Balding–Nichols · Bates · Beta (Generalized) · Beta rectangular · Continuous Bernoulli · Irwin–Hall · Kumaraswamy · Logit-normal · Noncentral beta · PERT · Raised cosine · Reciprocal · Triangular · U-quadratic · Uniform · Wigner semicircle
	supported on a semi-infinite interval	Benini · Benktander 1st kind · Benktander 2nd kind · Beta prime · Burr · Chi · Chi-squared (Noncentral · Inverse (Scaled)) · Dagum · Davis · Erlang (Hyper) · Exponential (Hyperexponential · Hypoexponential · Logarithmic) · *F* (Noncentral) · Folded normal · Fréchet · Gamma (Generalized · Inverse) · gamma/Gompertz · Gompertz (Shifted) · Half-logistic · Half-normal · Hotelling's *T*-squared · Inverse Gaussian (Generalized) · Kolmogorov · Lévy · Log-Cauchy · Log-Laplace · Log-logistic · Log-normal · Log-t · Lomax · Matrix-exponential · Maxwell–Boltzmann · Maxwell–Jüttner · Mittag-Leffler · Nakagami · Pareto · Phase-type · Poly-Weibull · Rayleigh · Relativistic Breit–Wigner · Rice · Truncated normal · type-2 Gumbel · Weibull (Discrete) · Wilks's lambda
	supported on the whole real line	Cauchy · Exponential power · Fisher's *z* · Kaniadakis κ-Gaussian · Gaussian *q* · Generalized normal · Generalized hyperbolic · Geometric stable · Gumbel · Holtsmark · Hyperbolic secant · Johnson's S_U · Landau · Laplace (Asymmetric) · Logistic · Noncentral *t* · Normal (Gaussian) · Normal-inverse Gaussian · Skew normal · Slash · Stable · Student's *t* · Tracy–Widom · Variance-gamma · Voigt
	with support whose type varies	Generalized chi-squared · Generalized extreme value · Generalized Pareto · Marchenko–Pastur · Kaniadakis *κ*-exponential · Kaniadakis *κ*-Gamma · Kaniadakis *κ*-Weibull · Kaniadakis *κ*-Logistic · Kaniadakis *κ*-Erlang · *q*-exponential · *q*-Gaussian · *q*-Weibull · Shifted log-logistic · Tukey lambda
Mixed univariate	continuous-discrete	Rectified Gaussian
Multivariate (joint)		*Discrete:* · Ewens · Multinomial (Dirichlet · Negative) · *Continuous:* · Dirichlet (Generalized) · Multivariate Laplace · Multivariate normal · Multivariate stable · Multivariate *t* · Normal-gamma (Inverse) · *Matrix-valued:* · LKJ · Matrix normal · Matrix *t* · Matrix gamma (Inverse) · Wishart (Normal · Inverse · Normal-inverse · Complex)
Directional		*Univariate (circular) directional*: Circular uniform · Univariate von Mises · Wrapped normal · Wrapped Cauchy · Wrapped exponential · Wrapped asymmetric Laplace · Wrapped Lévy · *Bivariate (spherical)*: Kent · *Bivariate (toroidal)*: Bivariate von Mises · *Multivariate*: von Mises–Fisher · Bingham
Degenerate and singular		*Degenerate*: Dirac delta function · *Singular*: Cantor
Families		Circular · Compound Poisson · Elliptical · Exponential · Natural exponential · Location–scale · Maximum entropy · Mixture · Pearson · Tweedie · Wrapped

Bernoulli Distribution

- Consider a biased coin with probability of heads p
- We can model the outcome of a single flip as a Bernoulli random variable
- We write $X \sim \text{Bernoulli}(p)$ and the PMF is

$$f_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

- Expectation? $\mathbb{E}[X] = p \cdot 1 + (1-p) \cdot 0 = p$
- Variance? $\text{Var}(X) = p \cdot 1^2 + (1-p) \cdot 0^2 - p^2 = p(1-p)$
- MGF of X? $M_X(t) = \mathbb{E}[e^{tX}] = pe^t + (1-p)e^0 \implies \mathbb{E}[X^r] = p, r \in \{1, 2, \dots\}$

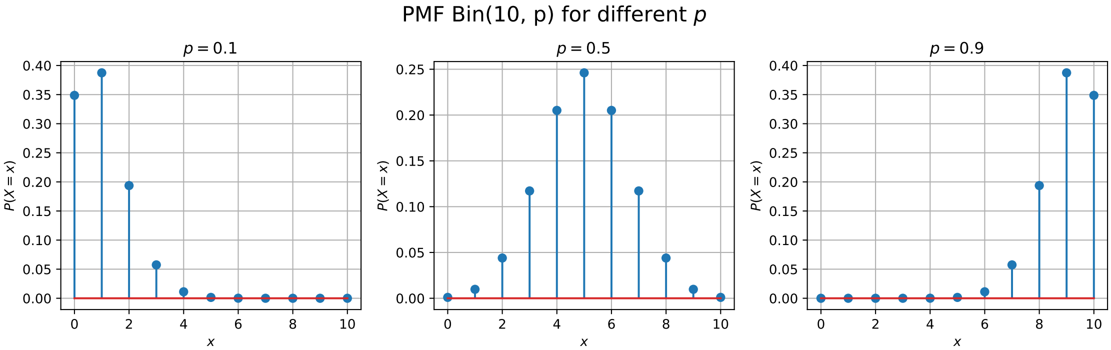
Binomial Distribution

- Consider n independent Bernoulli trials each with probability of success p
- Let X be the number of successes in these n trials
- We write $X \sim \text{Binomial}(n, p)$ and the PMF is

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

- Recall $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ is the number of ways to choose x elements from n without order

Binomial Distribution



Binomial Distribution

- Recall the **binomial expansion** $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$
- MGF?

$$
\begin{aligned}
M_X(t) = \mathbb{E}[e^{tX}] &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\
&= (pe^t + 1 - p)^n
\end{aligned}
$$

- $\mathbb{E}[X] = np$, $\mathbb{E}[X^2] = np(1-p) + n^2p^2$, $\mathsf{Var}(X) = np(1-p)$

For Later

- Take the MGF of a Binomial random variable:

$$M_X(t) = (pe^t + 1 - p)^n$$

- Consider the limit as $n \to \infty$ and $p \to 0$ such that $np = \lambda > 0$ is fixed:

$$\lim_{n\to\infty, p\to 0^+, np=\lambda} (pe^t + 1 - p)^n = \lim_{n\to\infty} \left(1 + \frac{\lambda(e^t - 1)}{n}\right)^n = e^{\lambda(e^t-1)}$$

Poisson Distribution

- Suppose we want to count a number of (successful) events in a fixed interval of time
- Assume that the events occurr independently with a constant rate λ per unit time
- Let X be the number of events in the interval of length 1
- We write $X \sim$ Poisson(λ) and the PMF is

$$f_X(x) = \frac{(\lambda)^x e^{-\lambda}}{x!}, \text{ for } x = 0, 1, 2, \ldots$$

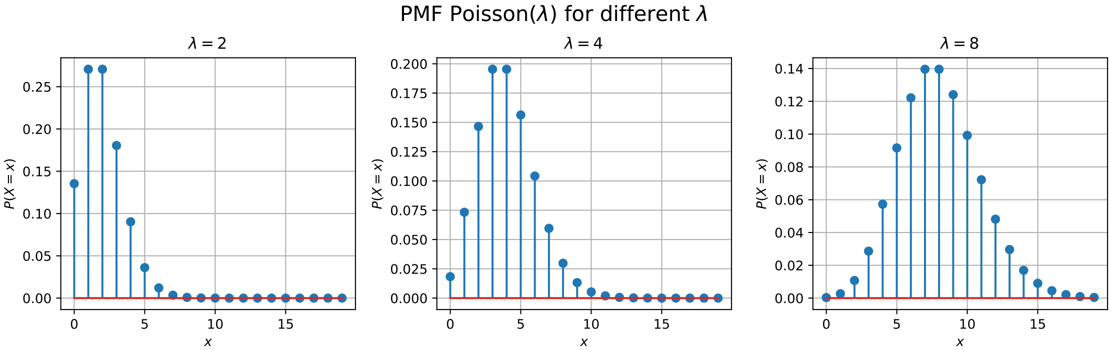
Poisson and Binomial

- MGF?

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \sum_{x=0}^{\infty} e^{tx} \frac{(\lambda)^x e^{-\lambda}}{x!} \\ &= \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x e^{-\lambda}}{x!} \\ &= e^{\lambda(e^t - 1)} \end{aligned}$$

- Recall that MGF uniquely determines the distribution of a random variable
- So Poisson is the limit of the Binomial as
  1. $n \to \infty$
  2. $p \to 0$
  3. $np = \lambda$ is fixed

Poisson Distribution



Geometric

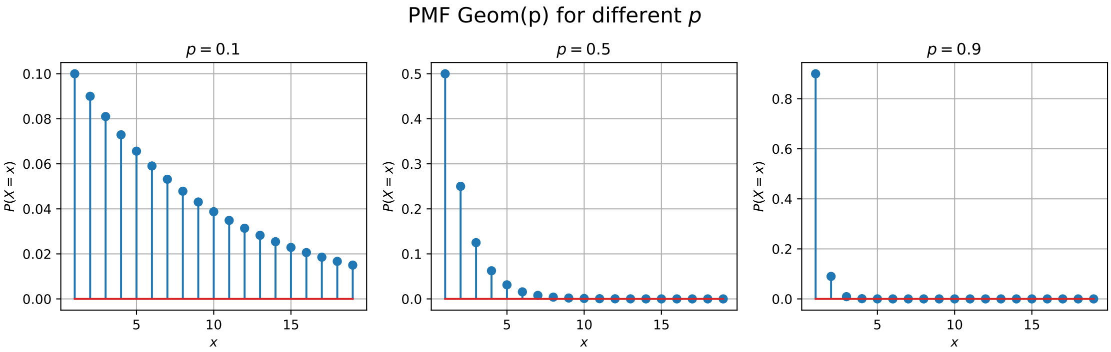
- Let X be the **number of trials** until the first success in a sequence of Bernoulli trials
- We write $X \sim$ Geometric(p) and the PMF is

$$f_X(x) = (1-p)^{x-1}p, \text{ for } x = 1, 2, \ldots$$

- MGF?

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \sum_{x=1}^{\infty} e^{tx}(1-p)^{x-1}p \\ &= pe^t \sum_{x=1}^{\infty} (e^t(1-p))^{x-1} \\ &= \frac{pe^t}{1-e^t(1-p)}, \text{ for } e^t(1-p) < 1 \iff t < -\log(1-p) \end{aligned}$$

Geometric Distribution



Can you guess the expectation of a Geometric random variable?
Suppose $p = 0.1$, how many times on average do you need to flip a coin until you get a head?

$$\mathbb{E}[X] = \frac{1}{p}$$

Negative Binomial

- Let X be the **number of trials** until the r**-th success** in a sequence of Bernoulli trials

- We write $X \sim \text{NegBin}(r, p)$ and the PMF is

$$f_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \text{ for } x = r, r+1, \ldots$$

- MGF?

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} p^r (1-p)^{x-r} \\ &= \sum_{x=r}^{\infty} \binom{x-1}{r-1} (pe^t)^r (1-p)^{x-r} \\ &= \left(\frac{pe^t}{1-(1-p)e^t} \right)^r \end{aligned}$$

Uniform Random Variable

- A random variable X is **uniformly distributed** on the interval $[a, b]$ if its PDF is

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

- The CDF is then

$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } x \in [a, b] \\ 1 & \text{if } x > b \end{cases}$$

- Expectation? $\mathbb{E}[X] = \frac{a+b}{2}$

- Variance? $\text{Var}(X) = \frac{(b-a)^2}{12}$

Exponential Distribution

- Let X be the time until the first event in a Poisson process with rate λ
- We write $X \sim \text{Exp}(\lambda)$ and the PDF is

$$f_X(x) = \lambda e^{-\lambda x}, \text{ for } x \geq 0$$

- The CDF is then

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- Expectation? $\mathbb{E}[X] = \frac{1}{\lambda}$
- MGF? $M_X(t) = \frac{\lambda}{\lambda - t}$, for $t < \lambda$

Memoryless Property

- The Exponential distribution is the only continuous RV with the **memoryless property**
- This means that for any $s, t \geq 0$, we have that

$$\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t)$$

- To see this, note that:

$$\begin{aligned}
\mathbb{P}(X > s + t | X > s) &= \frac{\mathbb{P}(X > s + t \cap X > s)}{\mathbb{P}(X > s)} \\
&= \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > s)} \\
&= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\
&= \mathbb{P}(X > t)
\end{aligned}$$

Weibull Distribution

- The **Weibull distribution** is a generalization of the Exponential distribution, $X \sim \text{Weibull}(\lambda, k), k > 0, \lambda > 0$

- The PDF is

$$f_X(x) = \begin{cases} \lambda k x^{k-1} e^{-\lambda x^k} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- The CDF is then

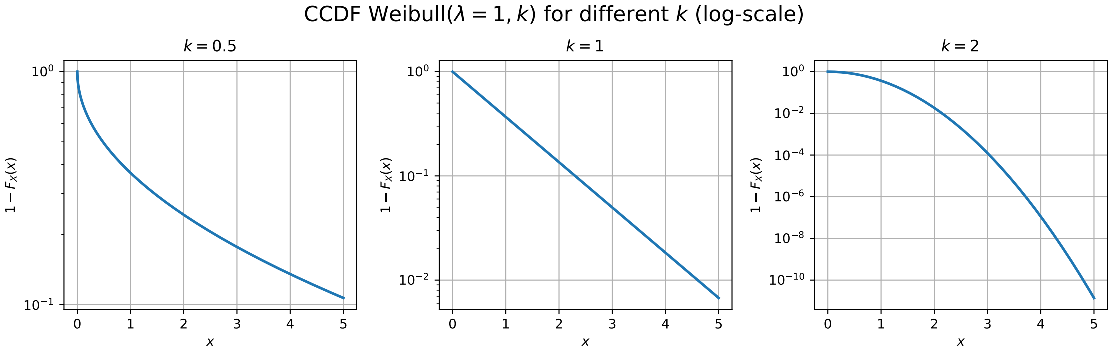
$$F_X(x) = \begin{cases} 1 - e^{-\lambda x^k} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- Memory?

$$\mathbb{P}(X > s + t | X > s) = \frac{e^{-\lambda (s+t)^k}}{e^{-\lambda s^k}}$$

- You can verify it is increasing (decreasing) in s if $k < 1$ ($k > 1$)

Different Weibull's



Gaussian Distribution

- The **Gaussian distribution** is also known as the **Normal distribution**, $X \sim N(\mu, \sigma^2)$
- The PDF is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- The CDF is then

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

- MGF? $M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$

Log-Normal Distribution

- The **Log-Normal distribution** is a distribution of a random variable whose logarithm is normally distributed

- Let X be a log-normal random variable, $Y = \log(X)$ is normally distributed

- The PDF of X is

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log(x)-\mu)^2}{2\sigma^2}}, \text{ for } x > 0$$

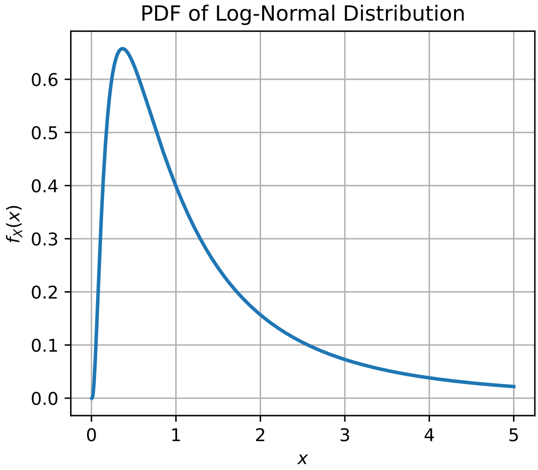
- The CDF is then

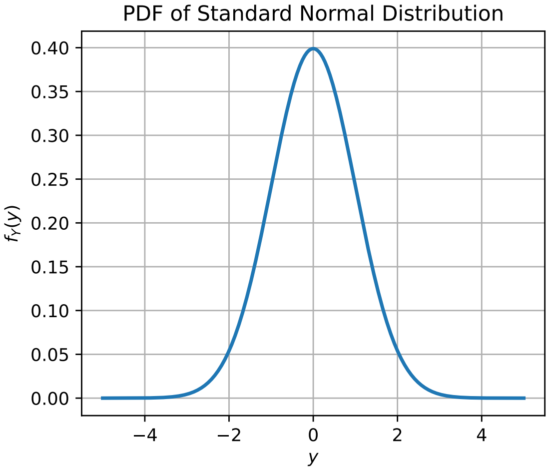
$$F_X(x) = \int_0^x \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\log(t)-\mu)^2}{2\sigma^2}} dt$$

- MGF? Does not exist for any t!

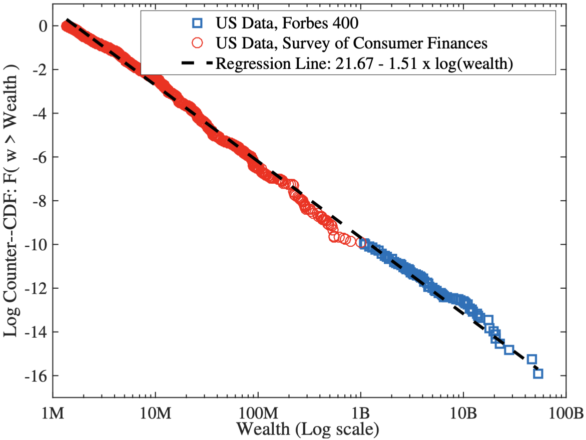
- Mean? $\mathbb{E}[X] = \mathbb{E}[e^{\log(X)}] = e^{\mu+\sigma^2/2}$

Log-Normal vs Normal





Motivating the Pareto



Also for income, consumption, city size, firm size, etc.

Deriving the Pareto

- We have seen that

$$\ln \mathbb{P}(X > x) \approx \ln(K) - \alpha \ln(x) \iff \mathbb{P}(X > x) \approx K x^{-\alpha}$$

- Differentiating we get the shape of the PDF:

$$f_X(x) = -\frac{d}{dx}\mathbb{P}(X > x) = \alpha K x^{-\alpha-1}$$

- For it to be integrable, cannot start at 0, so we add a lower bound x_m: $\int_{x_m}^{\infty} \alpha K x^{-\alpha-1} = K x_m^{-\alpha}$

- So the PDF of a Pareto is

$$f_X(x) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}} & \text{if } x \geq x_m > 0 \\ 0 & \text{otherwise} \end{cases}$$

Functions of Random Variables

- Let X be a random variable and $g : \mathbb{R} \to \mathbb{R}$ a "nice" function
- Define $Y = g(X)$
- What can we say about the **distribution of Y?**
- Recall that the CDF uniquely determines the distribution of a random variable

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y)$$

- But in general $\mathbb{P}(g(X) \leq y) \neq \mathbb{P}(X \leq g^{-1}(y))$
- For instance, $g(x) = x^2 \implies$ need a more general approach

Inverse Image

- Let $g : \mathbb{R} \to \mathbb{R}$ be a function
- The **inverse image** of a set $B \subset \mathbb{R}$ is defined as

$$g^{-1}(B) = \{x \in \mathbb{R} : g(x) \in B\}$$

- **Example:** Let $g(x) = x^2$ and $B = [0, 1]$, then $g^{-1}(B) = [-1, 1]$
- More generally:

$$\mathbb{P}(Y \in B) = \mathbb{P}(g(X) \in B) = \mathbb{P}(X \in g^{-1}(B))$$

Back to the CDF

$$
\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) \\
&= \mathbb{P}(X \in g^{-1}((-\infty, y])) \\
&= \int_{x \in g^{-1}((-\infty, y])} dF_X(x) \\
&= \begin{cases} \sum_{x \in g^{-1}((-\infty, y])} f_X(x) & \text{if } X \text{ is discrete} \\ \int_{x \in g^{-1}((-\infty, y])} f_X(x) dx & \text{if } X \text{ is continuous} \end{cases}
\end{aligned}
$$

Example: Number of Failures in n Trials

- We know that if X is the number of successes in n trials, then $X \sim \text{Binomial}(n, p)$
- By symmetry, we would expect that the number of failures $Y \sim \text{Binomial}(n, 1-p)$
- Let's prove this using the CDF:

$$
\begin{aligned}
F_Y(y) &= \sum_{\{x : g(x) \leq y\}} f_X(x) \\
&= \sum_{\{x : n-x \leq y\}} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=n-y}^{n} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=0}^{y} \binom{n}{n-x} p^{n-x} (1-p)^x, \quad \binom{n}{x} = \binom{n}{n-x}
\end{aligned}
$$

Example: Square of a Continuous RV

- Let X be a continuous random variable with CDF $F_X(x)$
- Let $Y = X^2$
- We want to find the CDF of Y

$$
\begin{aligned}
F_Y(y) &= \int_{x \in g^{-1}((-\infty, y])} f_X(x)dx \\
&= \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y}), \text{ for } y \geq 0
\end{aligned}
$$

- We can get the PDF of Y by differentiating $F_Y(y)$:

$$
f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} \left[f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right] & \text{for } y > 0 \\ 0 & \text{otherwise} \end{cases}
$$

Monotone Transformations

- We said that in general $\mathbb{P}(g(X) \leq y) \neq \mathbb{P}(X \leq g^{-1}(y))$
- But when $g : \mathbb{R} \to \mathbb{R}$ is a **strictly monotonic** function, we have that g^{-1} is well-defined
- In particular:

$$g^{-1}((-\infty, y]) = \{x \in \mathbb{R} : g(x) \leq y\} = \begin{cases} (-\infty, g^{-1}(y)] & \text{if } g \text{ is increasing} \\ [g^{-1}(y), \infty) & \text{if } g \text{ is decreasing} \end{cases}$$

Monotone Transformations

- Now we can write the CDF of $Y = g(X)$ as

$$
\begin{aligned}
F_Y(y) = \mathbb{P}(Y \leq y) &= \mathbb{P}(g(X) \leq y) \\
&= \begin{cases} \mathbb{P}(X \leq g^{-1}(y)) & \text{if } g \text{ is increasing} \\ \mathbb{P}(X \geq g^{-1}(y)) & \text{if } g \text{ is decreasing} \end{cases} \\
&= \begin{cases} F_X(g^{-1}(y)) & \text{if } g \text{ is increasing} \\ 1 - F_X(g^{-1}(y)^-) & \text{if } g \text{ is decreasing} \end{cases}
\end{aligned}
$$

1

[1] Recall $f(x^-) = \lim_{h \to 0^+} f(x - h)$

The Change of Variables Formula

- If X is **continous**, we have a general formula for the PDF of $Y = g(X)$:

$$
\begin{aligned}
f_Y(y) &= \begin{cases} \frac{d}{dy} F_X(g^{-1}(y)) & \text{if } g \text{ is increasing} \\ \frac{d}{dy}(1 - F_X(g^{-1}(y)^-)) & \text{if } g \text{ is decreasing} \end{cases} \\
&= \begin{cases} f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) & \text{if } g \text{ is increasing} \\ -f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) & \text{if } g \text{ is decreasing} \end{cases} \\
&= f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|
\end{aligned}
$$

- So if g is **strictly monotonic** and X is **continuous**, we can know the PDF immediately

Location-Scale Transformations

- A **location-scale transformation** is a transformation of the form $Y = \mu + \sigma X$
- We can write the CDF of Y as

$$
\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(\mu + \sigma X \leq y) \\
&= \mathbb{P}\left(X \leq \frac{y-\mu}{\sigma}\right) \\
&= F_X\left(\frac{y-\mu}{\sigma}\right)
\end{aligned}
$$

- The PDF of Y is then

$$
f_Y(y) = f_X\left(\frac{y-\mu}{\sigma}\right)\left|\frac{1}{\sigma}\right|
$$

- This should look familiar!

Example: Log-Normal Distribution

- Let $X \sim \mathcal{N}(\mu, \sigma^2)$
- Let $Y = e^X$
- We know that Y is log-normally distributed, this is how we can prove it:

$$
\begin{aligned}
f_Y(y) &= f_X(\log(y)) \left| \frac{1}{y} \right| \\
&= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\log(y)-\mu)^2}{2\sigma^2}} \frac{1}{y} \\
&= \frac{1}{y\sqrt{2\pi}\sigma} e^{-\frac{(\log(y)-\mu)^2}{2\sigma^2}}
\end{aligned}
$$