A.1 Probability Review - Outline

Discrete Distributions

X is a *discrete random variable* if it has a *probability mass function* p_k,

$$p_k = \mathrm{P}[X = k]$$
$$0 \leq p_k \leq 1$$
$$\sum_k p_k = 1$$

Example: If X is a Poisson random variable with mean λ then

$$p_k = e^{-\lambda}\frac{\lambda^k}{k!} \quad k = 0, 1, 2, \ldots$$

For us, discrete random variables will usually be counting / frequency variables, meaning the possible values are $\{0, 1, 2, \ldots\}$

Some key discrete distributions are included in a PDF handout on the exam.

Empirical Distribution

Suppose we observe 5 losses with amounts of:

$$500 \quad 600 \quad 650 \quad 650 \quad 750$$

If X is distributed according to the *empirical distribution* then

$$\mathrm{P}[X = 500] = \frac{1}{5} \quad \mathrm{P}[X = 650] = \frac{2}{5}$$

More generally, if we have n data points,

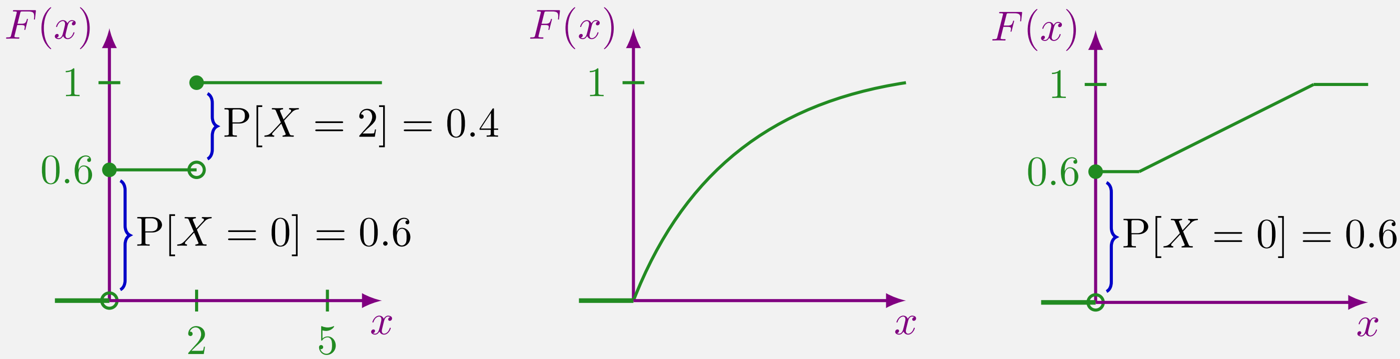
$$\mathrm{P}[X = x] = \frac{\text{\# of data points} = x}{\text{total \# of data points}} = \frac{\text{\# of data points} = x}{n}$$

Cumulative Distribution Function (CDF)

The cumulative distribution function (CDF) of X is

$$F(x) = \mathrm{P}[X \leq x]$$
$$F(\infty) = \lim_{x \to \infty} \mathrm{P}[X \leq x] = 1$$
$$F(-\infty) = 0$$

X is a *loss variable* if $X \geq 0$, in which case $F(x) = 0$ for all $x < 0$.



Discrete Continuous Mixed

Continuous Distributions

X is a continuous variable if $F(x)$ is continuous. For us, this will also imply that $F(x)$ is differentiable. The density $f(x)$ is given by

$$\begin{aligned} f(x) &= F'(x) \\ \mathrm{P}[a < X \leq b] &= \mathrm{P}[X \leq b] - \mathrm{P}[X \leq a] = F(b) - F(a) \\ &= \int_a^b f(x)dx \\ 1 &= \int_{-\infty}^{\infty} f(x)dx \\ f(x)\,dx &\text{ ``=''}\ \mathrm{P}[x < X \leq x + dx] \\ 0 &\leq f(x) \end{aligned}$$

There is no upper limit on $f(x)$. In particular, $f(x)$ is not a probability, and can be greater than 1. For example, if X is uniform on $(0, 0.1)$ then $f(x) = 1/(0.1 - 0) = 10$

Survival Function

$$\begin{aligned} S(x) &= \mathrm{P}[X > x] = 1 - F(x) = \text{survival function} \\ S(x) &= e^{-H(x)} \\ H(x) &= -\ln[S(x)] = \text{cumulative hazard function} \\ h(x) &= H'(x) = \text{hazard rate} \\ &= \frac{-S'(x)}{S(x)} = \frac{f(x)}{S(x)} \\ h(x)\,dx &= \frac{f(x)\,dx}{S(x)} \\ &\text{``=''}\ \mathrm{P}[x < X \leq x + dx \mid X > x] \end{aligned}$$

Exam Tables include $F(x)$ and $f(x)$ for many key distributions, from which you can find $h(x), H(x)$, and $S(x)$ if needed.

Percentiles

If X is continuous, then $100p^{th}$ percentile $= \pi_p(X)$
i.e., $F(\pi_p) = \mathrm{P}[X \le \pi_p] = p = p \times 100\%$
$\pi_p(X)$ is denoted $\mathrm{VaR}_p(X)$ on tables.

If X is discrete or mixed, the CDF may jump over p or $F(x) = p$ may not have a unique solution.

In words, if $F(x)$ jumps over p, then π_p is where the jump occurs.
If $F(x) = p$ doesn't have a unique solution, then all x-values with $F(x) = p$ are $100p^{th}$ percentiles, as is the right hand end point of that set.
In equations, π_p is a $100p^{th}$ percentile if

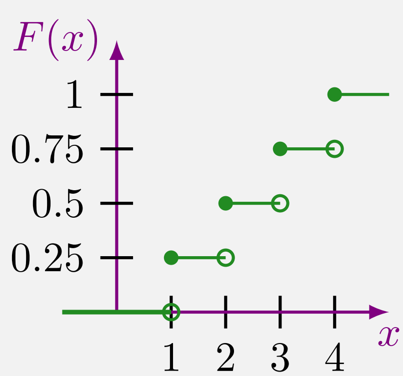
$$\mathrm{P}[X < \pi_p] \le p \le \mathrm{P}[X \le \pi_p]$$
$$\lim_{x \uparrow \pi_p} F(x) = F(\pi_p^-) \le p \le F(\pi_p)$$

This definition has never been tested.

Example

Let X be uniform on $\{1, 2, 3, 4\}$. Find the 10th, 30th, 40th, and 50th percentiles.

$$F(x) = \begin{cases} 0 & x < 1 \\ 0.25 & 1 \le x < 2 \\ 0.5 & 2 \le x < 3 \\ 0.75 & 3 \le x < 4 \\ 1 & 4 \le x \end{cases}$$



$F(x)$ jumps from 0 to 0.25 at 1, so $\pi_{0.1}(X) = 1$.
$F(x)$ jumps from 0.25 to 0.5 at 2, so $\pi_{0.3}(X) = \pi_{0.4}(X) = 2$.
$F(x) = 0.5$ for $2 \le x < 3$.
$F(x^-) \le 0.5$ for $x \le 3$, and $F(x) \ge 0.5$ for $x \ge 2$, so every point with $2 \le x \le 3$ is a 50th percentile.

Example

Suppose that $f(x) = cx^2$ for $0 < x < 5$ and is 0 otherwise. Find the 90th percentile of X.

$$
\begin{aligned}
1 &= \int_0^5 cx^2\,dx \\
&= c \cdot \frac{x^3}{3}\bigg|_0^5 = c \cdot \frac{125}{3} \\
c &= \frac{3}{125} = \frac{1}{\int\limits_0^5 x^2\,dx} \\
0.90 &= \int_0^{\pi_{0.9}} \frac{3}{125} x^2\,dx \\
0.9 &= \frac{1}{125}\pi_{0.9}^3 \\
\pi_{0.9} &= \boxed{4.83}
\end{aligned}
$$

Exercise 1

The cdf of X is $F(x) = 1 - \left(\frac{5}{x+5}\right)^3$ for $x > 0$. Find the hazard rate of x for $x > 0$.

Exercise 1

The cdf of X is $F(x) = 1 - \left(\frac{5}{x+5}\right)^3$ for $x > 0$. Find the hazard rate of x for $x > 0$.

$$\begin{aligned} S(x) &= \left(\frac{5}{x+5}\right)^3 \\ H(x) &= -\ln[S(x)] \\ &= -\ln\left[\left(\frac{5}{x+5}\right)^3\right] \\ &= -3\ln(5) + 3\ln(x+5) \\ h(x) &= 0 + \frac{3}{x+5} \end{aligned}$$

Remark: X is listed on the exam tables as a Pareto distribution with $\theta = 5$ and $\alpha = 3$. We will talk about Paretos more in A.2

Exercise 2

Find the density of X if the hazard rate is

$$h(x) = \begin{cases} 0 & x < 1 \\ 3 & 1 < x < 5 \\ 2x & 5 < x \end{cases}$$

Exercise 2

Find the density of X if the hazard rate is

$$h(x) = \begin{cases} 0 & x < 1 \\ 3 & 1 < x < 5 \\ 2x & 5 < x \end{cases}$$

$$\begin{aligned}
H(x) &= 0 \quad x < 1 \\
H(x) &= \int_1^x h(t)\,dt = \int_1^x 3\,dt = 3(x-1) \quad 1 < x < 5 \\
H(x) &= \int_{-\infty}^x h(t)\,dt \quad 5 < x \\
&= H(5) + \int_5^x 2t\,dt \quad 5 < x \\
&= 3 \cdot (5-1) + (x^2 - 25) \quad 5 < x \\
&= x^2 - 13 \quad 5 < x
\end{aligned}$$

Exercise 2 (Continued)

$$\begin{aligned}
H(x) &= \begin{cases} 0 & x < 1 \\ 3x - 3 & 1 < x < 5 \\ x^2 - 13 & 5 < x \end{cases} \\
S(x) &= e^{-H(x)} \\
&= \begin{cases} 1 & x < 1 \\ e^{-3x+3} & 1 < x < 5 \\ e^{-x^2+13} & 5 < x \end{cases} \\
f(x) &= -S'(x) \\
&= \begin{cases} 0 & x < 1 \\ 3e^{-3x+3} & 1 < x < 5 \\ 2xe^{-x^2+13} & 5 < x \end{cases}
\end{aligned}$$