# Probability: Describing Distributions

- Discrete Probability
  - Probability mass function
  - Empirical distribution
- CDF
- Continuous Distribution
  - Density
  - Survival Function
  - Hazard rate
  - Cumulative hazard rate function
  - Percentiles
- Examples

# Discrete Distributions

Probability mass function $p(x)$

$$p(x) = P[X = x]$$

$$0 \le p(x) \le 1$$

$$\sum_x p(x) = 1$$

Example: Poisson distribution

$$p(x) = e^{-\theta} \frac{\theta^x}{x!} \qquad x = 0, 1, 2, \ldots$$

# Empirical distribution

Collect some data,

e.g., have 5 claims

Amounts: 500 600 650 650 750

$X \sim$ empirical distribution

$$P[X=500] = \frac{1}{5}$$

$$P[X=650] = \frac{2}{5}$$

More generally, if we have $n$ claims

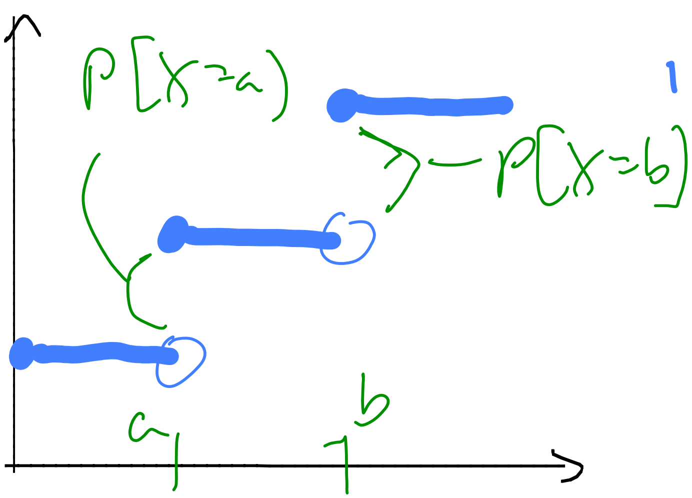
$$P[X=x] = \frac{\#\text{ of claims } = x}{n}$$

# Cumulative Distribution Function (CDF)

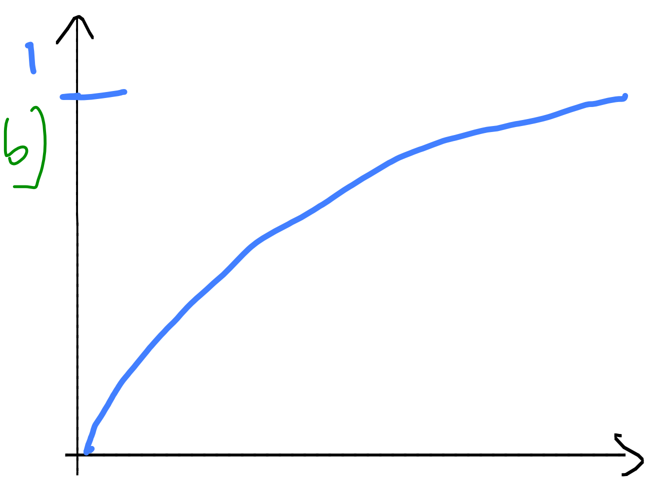
$$F(x) = P[X \le x]$$

$$F(\infty) = \lim_{x \to \infty} F(x) = 1, \quad \lim_{x \to -\infty} F(x) = 0$$

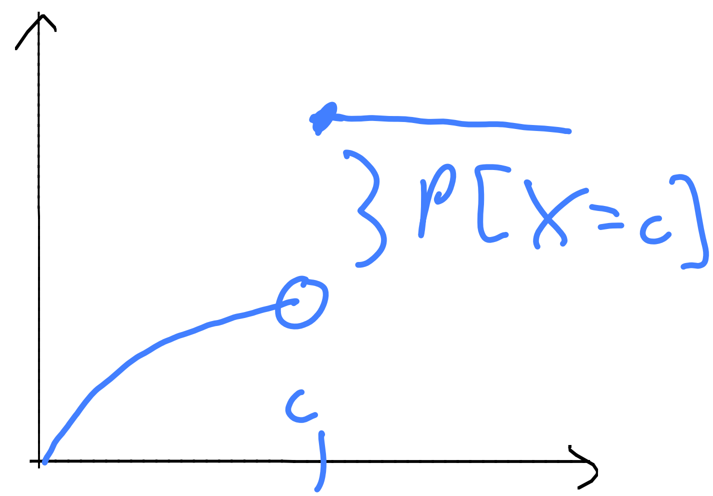
Usually, $F(x) = 0$ for all $x < 0$ for us.




Discrete




Continuous




Mixed

# Continuous Distributions

density $f(x) = F'(x)$

$$\mathbb{P}[a < X < b] = \int_a^b f(x)\,dx$$

so $0 \le f(x)$, $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$

---

Survival Function

$$S(x) = \mathbb{P}[X > x] = 1 - F(x)$$

Hazard rate: $h(x) = \dfrac{f(x)}{S(x)}$

Intuition:

$$f(x)\,dx \text{ "="} P[x < X \le x+dx]$$

$$dx\,h(x) = \frac{f(x)}{S(x)}\,dx = P[x < X \le x+dx \mid X > x]$$

i.e. $h(x)\,dx$ "=" Prob[fail just after time $x$ given alive at $x$]

Useful Formulas

$$h(x) = \frac{f(x)}{S(x)} = \frac{-S'(x)}{S(x)}$$

$$\int_0^x h(t)\,dt = -\int_0^x \frac{S'(t)}{S(t)}\,dt = -\ln S(x) + \ln S(0)$$

If $X \ge 0$, $S(0)=1$, get $= -\ln S(x) + 0$

$$H(x) = \int_0^x h(t)\,dt = -\ln S(x)$$

$$e^{-\int_0^x h(t)\,dt} = S(x)$$

$$\parallel$$

$$e^{-H(x)} = S(x),$$

$H(x)$ = cumulative hazard rate function

---

Mode: discrete case: $x$ maximizes $p(x)$

continuous: $x$ maximizes $f(x)$

# Percentiles

$100p^{th}$ percentile $\pi_p$

"Ideal" definition: $\pi_p$ is a $100p^{th}$ percentile if $F(\pi_p) = p$

Problem: 
- CDF may jump over $p$
- May not be unique

Actual definition: $\pi_p$ is a $100p^{th}$ percentile if:

$$F(\pi_p^-) \leq p \leq F(\pi_p)$$

$$\lim_{x \uparrow \pi_p} F(x) = P[X < \pi_p] \leq p \leq P[X \leq \pi_p]$$

[3-CAS.F03.17] Losses have an Inverse Exponential distribution. The mode is 10,000. Calculate the median. $F(x) = \frac{1}{2}$

A. Less than 10,000
B. At least 10,000, but less than 15,000
C. At least 15,000, but less than 20,000
D. At least 20,000, but less than 25,000
$\boxed{\text{E.}}$ At least 25,000

$\text{Mode} = \frac{\theta}{2}$ , $F(x) = e^{-\frac{\theta}{x}}$

$10{,}000 = \frac{\theta}{2}$

$20{,}000 = \theta$

$\frac{1}{2} = e^{-\frac{20{,}000}{x}}$

$\ln\left(\frac{1}{2}\right) = -\frac{20{,}000}{x}$

$x = 28{,}854$

[3-CAS.F03.19] For a loss distribution where $x \geq 2$, you are given:

$f(x) = 0 \quad x < 2$

$h(x) = 0 \quad x < 2$

(i) The hazard rate function: $h(x) = \dfrac{z^2}{2x}$, for $x \geq 2$

(ii) A value of the distribution function: $F(5) = 0.84$.

Calculate $z$.

$\boxed{\text{A. } 2}$ B. 3 C. 4 D. 5 E. 6

$S(x) = e^{-H(x)}$

$S(5) = 1 - F(5) = .16 = e^{-H(5)}$

$H(5) = -\ln(.16) = \dfrac{z^2}{2} \ln(2.5)$

$\boxed{z = 2}$

$H(5) = \int_0^5 h(x)\,dx$

$= \int_0^2 0\,dx + \int_2^5 \dfrac{z^2}{2x}\,dx = \dfrac{z^2}{2} \ln x \Big|_2^5 = \dfrac{z^2}{2} \ln(2.5)$