M339J: March 1st, 2021.

Percentiles.

Def'n. The $100p^{th}$ percentile of a random variable X is any value π_p such that

$$\boxed{F_X(\pi_p -) \leq p \leq F_X(\pi_p)}.$$

In particular, the 50^{th} percentile is called the median.

Special Case: Continuous dist'ns w/ a strictly positive density.

If we have $f_X(x) > 0$,
then $F_X(a) = \int_{-\infty}^{a} f_X(x)\,dx$ is strictly increasing.

$\Rightarrow$ F_X is one-to-one

$\Rightarrow$ F_X has an inverse

$\Rightarrow$ $F_X(\pi_p) = p \Leftrightarrow \pi_p = F_X^{-1}(p)$.

Problem. Find the ratio of the 90^{th} percentile to the median of an exponential dist'n w/ mean 10.

→: The cdf of $X \sim$ Exponential(mean = 10) is

$$F_X(x) = 1 - e^{-\frac{x}{10}} \quad \text{for } x > 0$$

Call the 90^{th} percentile a;
call the median, i.e., the 50^{th} percentile b.
Need: a/b.

Let $p \in (0,1)$. We want to find an expression for the $100p^{th}$ percentile of an exponential.

$$F_X(\pi_p) = p$$
$$1 - e^{-\frac{\pi_p}{\theta}} = p$$
$$1 - p = e^{-\frac{\pi_p}{\theta}} \quad / \ln$$
$$\ln(1-p) = -\frac{\pi_p}{\theta}$$
$$\boxed{\pi_p = -\theta \cdot \ln(1-p) = F_X^{-1}(p)}$$

$$\Rightarrow \left.\begin{array}{l} a = -\theta \cdot \ln(1-0.9) = -\theta \cdot \ln(0.1) \\ b = -\theta \cdot \ln(1-0.5) = -\theta \cdot \ln(0.5) \end{array}\right\}$$

$$\Rightarrow \quad \frac{a}{b} = \frac{\ln(0.1)}{\ln(0.5)} = 3.32$$

$\blacksquare$

Generating functions.

Def'n. Let X be a random variable.

- The moment generating function (mgf) is denoted by M_X and defined by
$$M_X(t) := \mathbb{E}\left[e^{t \cdot X}\right]$$
for all $t \in \mathbb{R}$ such that the expectation exists.

Note: At least $t=0$ is in the domain of M_X.

- The probability generating function (pgf) is denoted by P_X and defined by
$$P_X(s) := \mathbb{E}\left[s^X\right]$$
for all $s > 0$ such that the expectation exists.

Note: At least $s = 1$ is in the domain of P_X.

Note:

$$s \longleftrightarrow e^t$$

$$P_X(s) = M_X(\ln(s))$$

and

$$M_X(t) = P_X(e^t)$$

Sums of Independent Random Variables.

Thm. Let $\{X_k \,;\, k = 1, 2, \ldots\}$ be independent random variables. Define their "running" sums by

$$S_n = X_1 + X_2 + \cdots + X_n \quad \text{for all } n \in \mathbb{N}.$$

Then,

$$M_{S_n}(t) = \prod_{k=1}^{n} M_{X_k}(t) \quad \text{for all } n$$

and

$$P_{S_n}(s) = \prod_{k=1}^{n} P_{X_k}(s) \quad \text{for all } n.$$

$\rightarrow$:

$$M_{S_n}(t) = \mathbb{E}\left[e^{t \cdot S_n}\right] = \mathbb{E}\left[e^{t(X_1 + X_2 + \cdots + X_n)}\right]$$

$$= \mathbb{E}\left[e^{t \cdot X_1} \cdot e^{t \cdot X_2} \cdot \cdots \cdot e^{t X_n}\right]$$

$$\ldots$$

independent

$$= \mathbb{E}\left[e^{t \cdot X_1}\right] \cdot \mathbb{E}\left[e^{t \cdot X_2}\right] \cdots \mathbb{E}\left[e^{t \cdot X_n}\right]$$

$$= M_{X_1}(t) \cdot M_{X_2}(t) \cdots M_{X_n}(t) \quad \blacksquare$$

Consequences:

I. Sums of independent normal r.v.s are normal.

II. Sums of independent gamma-distributed w/ the same parameter θ are themselves gamma-distributed.

III. Sums of independent Poisson r.v.s are Poisson.

We can also easily identify the parameter values!