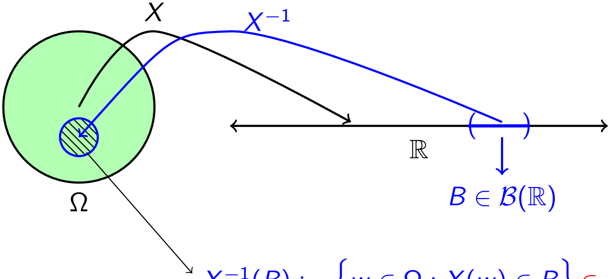
MA 6.101
Probability and Statistics

Tejas Bodas

Assistant Professor, IIIT Hyderabad

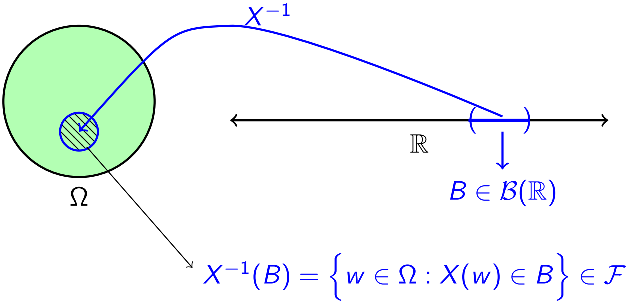
Random variables ($\Omega' = \mathbb{R}$)



$$X^{-1}(B) := \Big\{ w \in \Omega : X(w) \in B \Big\} \in \mathcal{F}$$

- $\Omega \xrightarrow{X} \mathbb{R}, \quad \mathcal{F} \xrightarrow{X} \mathcal{B}(\mathbb{R})$, and $P(.) \xrightarrow{X} P_X(.)$
- Care must be taken such that the events you consider in the new event space $\mathcal{B}(\mathbb{R})$ are also valid events included in $\mathcal{F}$.
- $X^{-1}(B)$ is called as the preimage or the inverse image of B.

Definition of a random variables



A random variable X is a map $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ such that for each $B \in \mathcal{B}(\mathbb{R})$, the inverse image $X^{-1}(B) := \{w \in \Omega : X(w) \in B\}$ satisfies

$$X^{-1}(B) \in \mathcal{F} \text{ and}$$
$$P_X(B) = \Pr(w \in \Omega : X(w) \in B)$$

Induced measure P_X and CDF

- The cumulative distribution function (CDF) $F_X(x)$ can be expressed using induced measure P_X.
- Since the domain of P_X is $\mathcal{B}(\mathbb{R})$, we have seen that $\mathcal{B}(\mathbb{R})$ is made up of sets of the form $(-\infty, a]$ for $a \in \mathbb{R}$.
- $P_X((-\infty, x]) = \mathbb{P}\{w \in \Omega : X(w) \leq x\} := F_X(x)$.
- This is a general definition of CDF (applicable for both continuous or discrete).
- If $F_X(\cdot)$ is continuous (resp. piecewise continuous), then X is continuous (resp. discrete) random variable.

For a r.v. X, its CDF satisfies the following

- $F_X(\infty) = 1$ and $F_X(-\infty) = 0$ when $P(-\infty < X < \infty) = 1$.
- $F_X : \mathbb{R} \to [0, 1]$ is non-decreasing and right continuous.
- At point of discontinuity x we have
  1. right hand limit $F_X(x+) := \lim_{\epsilon \downarrow 0} F_X(x + \epsilon)$
  2. left hand limit $F_X(x-) := \lim_{\epsilon \uparrow 0} F_X(x - \epsilon)$
  3. $F_X(x-) \neq F_X(x+)$.
  4. $F_X(x)$ could be set to either of the two. Which one?
- Right continuity mandates that at point of discontinuity, we have $F_X(x) = F_X(x+)$.
- By default, $F_X(x) = F_X(x+) = F_X(x-)$ if $F_X(x)$ is continuous at x.

Right-continuity

$F_X : \mathbb{R} \to [0, 1]$ is non-decreasing and right continuous.

Proof

- Consider $a < b$ where a and b are arbitrary. We want to show that $F_X(a) \le F_X(b)$.
- Define $A := \{\omega \in \Omega : X(\omega) \le a\}$, $B := \{\omega \in \Omega : X(\omega) \le b\}$.
- Easy to see that $A \subseteq B$ and hence $\mathbb{P}(A) \le \mathbb{P}(B)$.
- $F_X(a) = P_X((-\infty, a]) = \mathbb{P}(A) \le \mathbb{P}(B) = F_X(b)$.
- This proves the non-decreasing part.

Right-continuity

$F_X : \mathbb{R} \to [0,1]$ is non-decreasing and right continuous.

Proof for right-continuity

- We want to prove that $F_X(x) = F_X(x+)$.
- Consider a sequence of numbers $\{x_n\}$ decreasing to x. In this case, we have $F_X(x+) = \lim_{x_n \downarrow x} F_X(x_n)$.
- Define $A_n := \{\omega : X(\omega) \leq x_n\}$ and $A := \{\omega : X(\omega) \leq x\}$.
- Is $A_n \uparrow A$ or $A_n \downarrow A$? Clearly, $A_n \downarrow A$.
- From continuity of probability, $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.
- This implies $\lim_{x_n \downarrow x} F_X(x_n) = F_X(x)$. $\square$
- You cannot prove the other way by considering $x_n \uparrow x$ because $\cup_n(-\infty, x_n] = (-\infty, x)$ and $P_X(-\infty, x) \neq F_X(x)$.

Continuous random variables

- If Ω' is countable, then the random variable is called a discrete random variable.
- In this case it is convenient to use $\mathcal{F}'$ as power-set.
- All the probability measure is concentrated at discrete points.
- If $\Omega' \subseteq \mathbb{R}$ or uncountable, then the random variable is a continuous random variable.
- In this case, $\mathcal{F}' = \mathcal{B}(\mathbb{R})$.
- Intuitively, in a continuous random variable, the unit probability measure is spread continuously (like spreading a fluid) over the range of the random variable.

Examples of Continuous random variables

- Pick a number uniformly from $[a, b]$.
- Time interval between successive customers entering DMart.
- Travel time from office to home.
- Level of water in a dam or pending workload on a server.

Continuous random variables

- A random variable X is continuous if there exists a non-negative real valued probability density function (PDF) $f_X(\cdot)$ such that $F_X(x) = \int_{u=-\infty}^{x} f_X(u)du$.
- $P_X(B) = \int_{u \in B} f_X(u)du$. $P_X(\mathbb{R}) = \int_{u=-\infty}^{\infty} f_X(u)du = 1$.
- $P_X(a \leq X \leq b) = \int_a^b f_X(u)du$. (Area under the curve)
- $P_X(a \leq X \leq b) = P_X(a < X < b) = P_X(a \leq X < b) = P_X(a < X \leq b)$
- $P_X(X = a) = 0$. (no mass at any point)

$\frac{dF_X(x)}{dx} = f_X(x)$ or $P_X(x < X \leq x + h) \simeq f_X(x)h$.

Mean, Variance, Moments

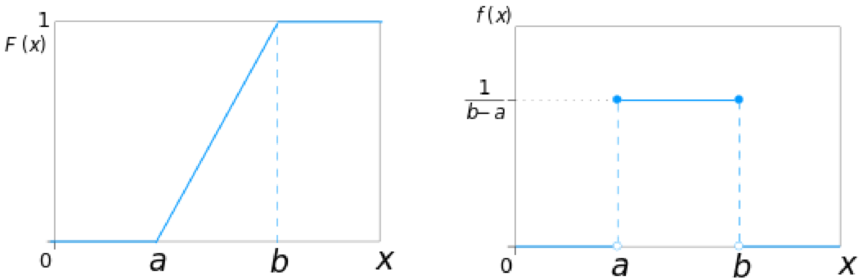
- $E[X] = \int_{-\infty}^{\infty} u f_X(u) du$
- $E[X^n] = \int_{-\infty}^{\infty} u^n f_X(u) du$
- $E[g(X)] = \int_{-\infty}^{\infty} g(u) f_X(u) du$
- $Var[X] = E[g(X)]$ where $g(x) = (x - E[X])^2$.
- For $Y = aX + b$, $E[Y] = aE[X] + b$.
- For $Y = aX + b$, $F_Y(y) = F_X(\frac{y-b}{a})$ when $a \geq 0$.
- For $Y = aX + b$ and $a < 0$, $F_Y(y) = 1 - F_X(\frac{y-b}{a})$.

Standard Examples

Uniform random variable ($U[a, b]$)

- This is a real valued r.v.
- Its pdf $f_X(x) = \frac{1}{b-a}$ for all $x \in [a, b]$.
- Its CDF is given by $F_X(x) = \begin{cases} 0 \text{ for } x < a. \\ \frac{x-a}{b-a} \text{ for } x \in [a, b] \\ 1 \text{ otherwise.} \end{cases}$
- HW: Verify $E[X] = \frac{a+b}{2}$ and $Var(X) = \frac{(b-a)^2}{12}$

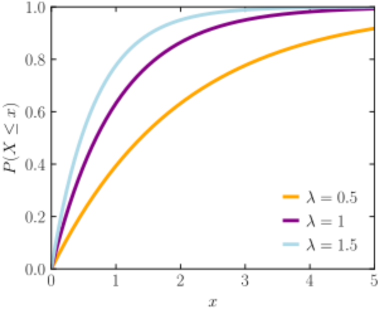
$U[a, b]$



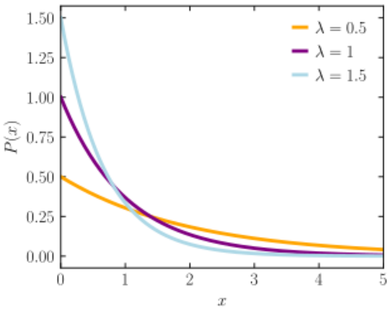
Exponential random variable ($Exp(\lambda)$)

- This is a non-negative r.v. with parameter λ.
- Its pdf $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.
- Its CDF is given by $F_X(x) = 1 - e^{-\lambda x}$ for $x \geq 0$.
- $E[X] = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$
- $E[X^n] = \frac{n!}{\lambda^n}$

$Exp(\lambda)$



1.0
0.8
0.6
0.4
0.2
0.0
$P(X \leq x)$
0
1
2
3
4
5
x
$\lambda = 0.5$
$\lambda = 1$
$\lambda = 1.5$



1.50
1.25
1.00
0.75
0.50
0.25
0.00
$P(x)$
0
1
2
3
4
5
x
$\lambda = 0.5$
$\lambda = 1$
$\lambda = 1.5$

Significance of Exponential r.v.

- Building blocks for Continuous time Markov Chains.
- Demonstrate memory-less property (to be seen formally soon).
- $P(X > a + h | X > a) = \frac{e^{-\lambda(a+h)}}{e^{-\lambda(a)}} = e^{-\lambda(h)} = P(X > h)$.
- Used extensively in Queueing theory to model inter-arrival time and service time of jobs.

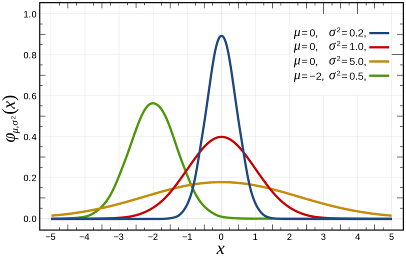
Recap

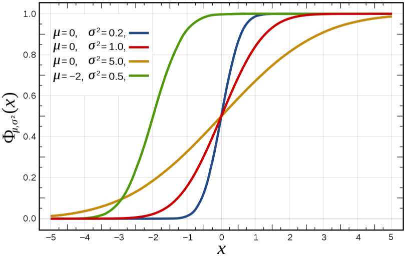
- Discrete random variables and relation between $\mathbb{P}, P_X, F_X, p_X$.
  - Relation between p_X and F_X
    $F_X(a) = \sum_{x \leq a} p_X(x) = \mathbb{P}\{\omega \in \Omega : X(\omega) \leq a\}$.
  - Relation between P_X and F_X
    $F_X(a) := P_X((-\infty, a]) = \mathbb{P}\{\omega \in \Omega : X(\omega) \in (-\infty, a]\}$
  - Relation between P_X and p_X
    $p_X(a) := P_X(\{a\}) = \mathbb{P}\{\omega \in \Omega : X(\omega) = a\}$
- Continuous variables and relation between $\mathbb{P}, P_X, F_X, f_X$
  - Relation between f_X and F_X is $F_X(a) = \int_{-\infty}^{a} f_X(x)dx$.
  - $\frac{dF_X(x)}{dx} = f_X(x)$ or $P_X(x < X \leq x + h) \simeq f_X(x)h$.
- Mean, Variance, Moments, $E[g(X)]$, Linearity & Examples

$F_X : \mathbb{R} \to [0, 1]$ is non-decreasing and right continuous.

Gaussian random variable ($\mathcal{N}(\mu, \sigma^2)$)

- This is a real valued r.v. with two paramerters, μ and σ.
- Its pdf $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ for all $x \in \mathbb{R}$.
- Verify: $\int_{-\infty}^{\infty} f_X(x)dx$ $E[X] = \mu$ and $Var(X) = \sigma^2$.





Standard Normal random variable ($\mathcal{N}(0,1)$)

- When $\mu = 0$ and $\sigma = 1$, it is called as a standard normal.
- In this case $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$.
- What is $\int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} dx$? How do you even solve this? $(= \sqrt{2\pi})$
- The CDF of standard normal, denoted by $\Phi(x)$ is given by
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$
- $Q(x) := 1 - \Phi(x)$ is the Complimentary CDF ($P(X > x)$). A closely related cousin in the error function $erf(x) = \frac{2}{\sqrt{pi}} \int_0^x e^{t^2} dt$.
- $\Phi =$These values are recorded in a table. (Fig. 3.10 in Bertsekas)
- `https://en.wikipedia.org/wiki/Gaussian_function`

Normality preserved under Linear Transformations

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Y = aX + b$ is also a normal variable with $E[Y] = a\mu + b$ and variance $a^2\sigma^2$. (To be proved later)

- Suppose X is standard normal, then find a and b such that $Y \sim \mathcal{N}(\mu, \sigma^2)$
- In this case, the CDF of Y in terms of X is given by $\Phi(\frac{x-\mu}{\sigma})$.

Significance of Gaussian r.v.

- Key role in Central limit theorem.
- $\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ where X_i is any random variable with mean μ and variance σ^2.
- Building block for multinomial Gaussian vector and Gaussian processes (GP).
- Gaussian process are used in Bayesian Optimization (black-box optimization).
- Brownian motion is a type of GP and is used in Finance.
- GP Regression, Gaussian mixture models, used widely in ML.

List of Probability distributions ...

https://en.wikipedia.org/wiki/List_of_probability_distributions

Important ones are Beta, Gamma, Erlang, Logistic, Weibull ....

Moment generating function

- The moment generating function (MGF) of a random variable X is a function $M_X : \mathbb{R} \to [0, \infty]$ defined by $M_X(t) = E[e^{tX}]$.
- If X is discrete, $M_X(t) = \sum_{x \in \Omega'} e^{tx} p_X(x)$.
- If X is continuous, $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$.
- For $Exp(\lambda)$ variable, $M_X(t) = \frac{\lambda}{\lambda - t}$ for $\lambda < t$.
- Define $D_X := \{t : M_X(t) < \infty\}$. D_X is called the region of convergence (ROC). $t = 0$ is always part of ROC.
- HW: Find the MGF for a random variable X that has the following distributions: Binomial(n,p), Normal $\mathcal{N}(0, 1)$, Poisson(λ)

MGF

- If $M_X(t)$ is finite for all $|t| \leq \epsilon$ and for some $\epsilon > 0$ then $M_X(t)$ is infinitely differentiable on $(-\epsilon, \epsilon)$. (Property without proof)
- Let $M_X{}^{(r)}(t) := \frac{d^r}{dt^r} M_X(t)$ (r^{th}-derivative of $M_X(t)$)
- It can be shown that $M_X{}^{(r)}(t) = E[e^{tX} X^r]$ for all r and $|t| \leq \epsilon$.
- $E[X^r] = M_X{}^{(r)}(0)$
- HW: Work out these things for $Exp(\lambda)$
- HW: Find MGF for all random variables studied till now