6.041/6.341 Quiz I Review (Fall 08)

I. We will begin our quiz review by summarizing the important concepts and formulae from the first part of this class. We will go over:

   – Sample space and probability,

   – Discrete random variables, and

   – Introduction to continuous random variables.

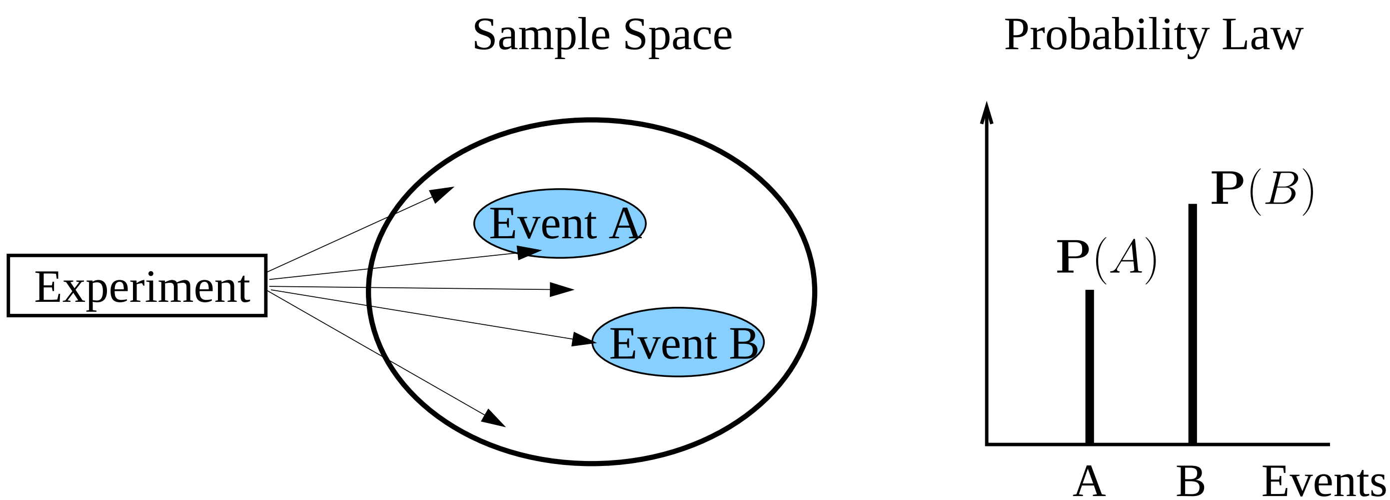
II. The second half of the review will consist of solving previous quizzes.

Probabilistic Models

Consist of:

1. The **sample space**, Ω: the set of all possible outcomes of an experiment.

2. The **probability law**: assigns to each **event**, which is a set A of possible outcomes, a nonnegative number $\mathbf{P}(A)$, called the probability of A.



Probability Axioms

1. **Nonnegativity**: $\mathbf{P}(A) \geq 0$, for every event A.
2. **Additivity**: If A and B are disjoint events, then

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B).$$

3. **Normalization**: $\mathbf{P}(\Omega) = 1$.

Some Properties of Probability Laws

These can be deduced from the axioms and verified using **Venn Diagrams.**

Let A, B and C be events.

1. If $A \subset B$, then $\mathbf{P}(A) \leq \mathbf{P}(B)$.
2. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.
3. $\mathbf{P}(A \cup B) \leq \mathbf{P}(A) + \mathbf{P}(B)$.
4. $\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(A^c \cap B) + \mathbf{P}(A^c \cap B^c \cap C)$.

Discrete Models

- **Discrete Probability Law**: If Ω is finite, then any event has the form $A = \{s_1, s_2, \ldots, s_n\}$, where $s_i \in \Omega$, and,

$$\mathbf{P}(A) = \mathbf{P}(s_1) + \mathbf{P}(s_2) + \cdots + \mathbf{P}(s_n).$$

- **Discrete Uniform Probability Law**: If Ω consists of n equally likely outcomes, then,

$$\mathbf{P}(A) = \frac{\text{number of elements of } A}{n}.$$

Conditional Probability

- Given B, with $\mathbf{P}(B) > 0$, the conditional probability of A is

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

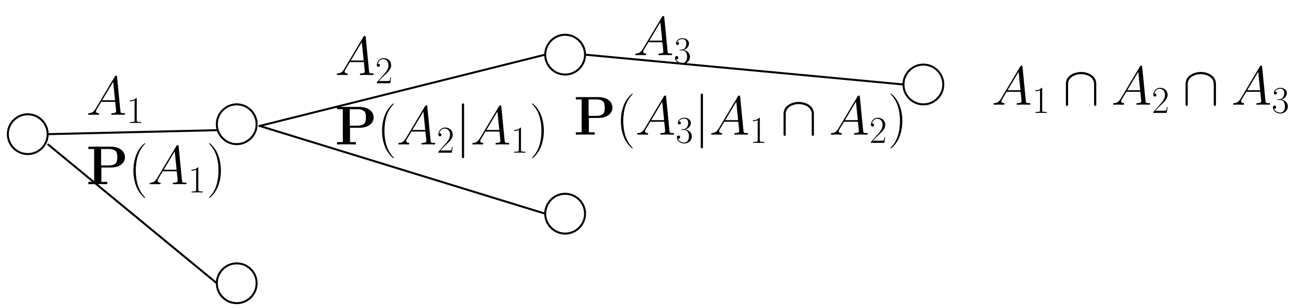
- $\mathbf{P}(A|B)$ is a new probability law on Ω and satisfies the probability axioms.
  1. $\mathbf{P}(A|B) \geq 0$.
  2. $\mathbf{P}(\Omega|B) = 1$.
  3. $\mathbf{P}(A_1 \cup A_2|B) = \mathbf{P}(A_1|B) + \mathbf{P}(A_2|B)$, when $A_1 \cap A_2 = \emptyset$.
- $\mathbf{P}(A|B)$ can be viewed as a probability law on a new universe, B, where all the conditional probability is concentrated.

Multiplication Rule

- For finding (unconditional) probabilities of events when an experiment has a sequential nature and conditional probabilities are known.
- Assuming all conditioning events have positive probabilities:

$$\mathbf{P}(\cap_{i=1}^{n} A_i) = \mathbf{P}(A_1)\mathbf{P}(A_2|A_1)\mathbf{P}(A_3|A_1 \cap A_2)\ldots\mathbf{P}(A_n|\cap_{i=1}^{n-1} A_i).$$

- The possible outcomes of the sequential experiment can be represented using a tree. The probability of the leaf (event) is the product of the probabilities along the traversed path.

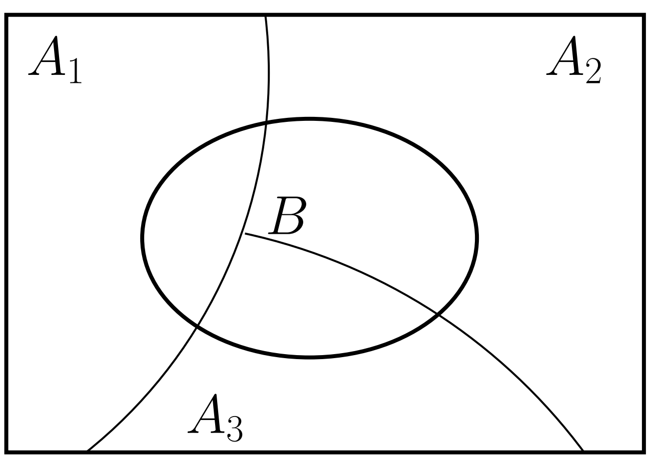


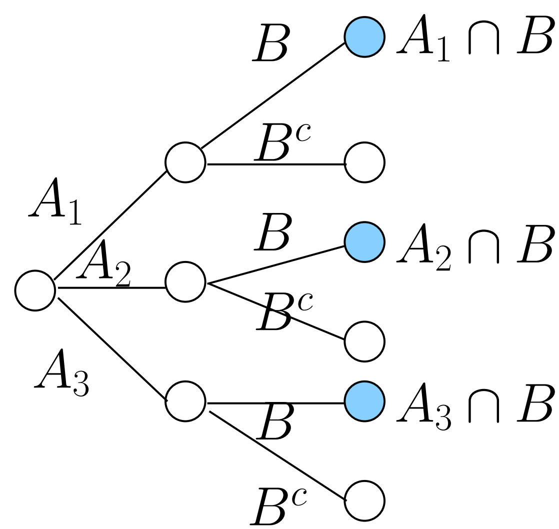
Total Probability Theorem

A divide-and-conquer approach to find the probability of event B.

Let $A_1, \ldots, A_n$ be disjoint events that partition Ω and $\mathbf{P}(A_i) > 0$, $\forall i$. Then for any B,

$$\mathbf{P}(B) = \sum_{i=1}^{n} \mathbf{P}(A_i \cap B) = \sum_{i=1}^{n} \mathbf{P}(B|A_i)\mathbf{P}(A_i).$$



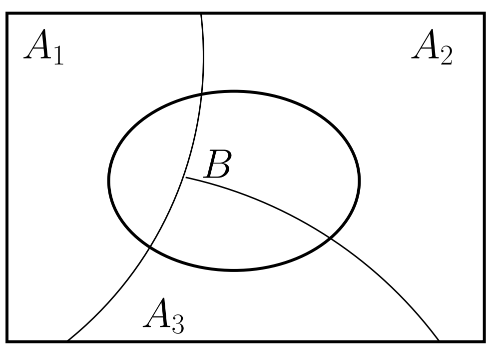


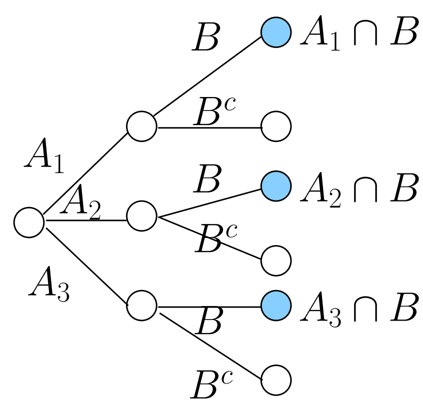
Bayes Rule

Used for **inference**. Events $A_1, \ldots, A_n$ are "causes" that result in event B. We observe B and want to infer the cause, i.e., we need the probability that the cause was A_i.

Let $A_1, \ldots, A_n$ be disjoint events that partition Ω and $\mathbf{P}(A_i) > 0$, $\forall i$. Then for B, where $\mathbf{P}(B) > 0$,

$$\mathbf{P}(A_i|B) = \frac{\mathbf{P}(B|A_i)\mathbf{P}(A_i)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B|A_i)\mathbf{P}(A_i)}{\sum_{i=1}^{n} \mathbf{P}(B|A_i)\mathbf{P}(A_i)}.$$





Independence

- A and B are
  - **independent** if $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$, or equivalently, $\mathbf{P}(A|B) = \mathbf{P}(A)$, if $\mathbf{P}(B) > 0$.
  - **conditionally independent** given C, where $\mathbf{P}(C) > 0$, if $\mathbf{P}(A \cap B|C) = \mathbf{P}(A|C)\mathbf{P}(B|C)$, or equivalently, $\mathbf{P}(A|B \cap C) = \mathbf{P}(A|C)$, if $\mathbf{P}(B \cap C) > 0$.
  - Note: Independence **does not imply** conditional independence and vice versa.
- $A_1, A_2, \ldots, A_n$ are
  - **independent** if $\mathbf{P}(\bigcap_{i \in S} A_i) = \prod_{i \in S} \mathbf{P}(A_i)$, for every subset $S \subset \{1, 2, \ldots, n\}$.
  - Note: Pairwise independence **does not imply** independence.

Counting

Can be used to find probabilities when the discrete uniform probability law applies.

- **The counting principle**: For an r stage process with n_i choices at stage i, the total number of choices is $n_1 n_2 \dots n_r$.
- $k-$**Permutations** of n objects (no replacement): the number of ways k objects can be picked from n and arranged in a sequence is $n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$.

  ("AB" and "BA" are counted as distinct objects.)
- **Combintations** of k objects out of n (no replacement): the number of ways k-element subsets can be chosen out of an n-element set is $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

  (No ordering of selected objects; $\{A, B\}$ is the same as $\{B, A\}$.)

- **Partitions**: The number of ways to partition an n-element set into r disjoint subsets, the ith subset containing n_i elements, is,

$$\begin{aligned}\binom{n}{n_1, n_2, \ldots n_r} &= \binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\cdots-n_{r-1}}{n_r} \\ &= \frac{n!}{n_1!n_2!\ldots n_r!}.\end{aligned}$$

Note: $\sum_{i=1}^{r} n_i = n$.

Discrete Random Variables

For any probabilistic model of an experiment,

- A **discrete random variable**, X,
  - is a *function* that maps the outcomes of the experiment to a set, S, containing finite or countably infinite real numbers.
  - has a **PMF** that gives the probability of each of the values in S.
- A **function of a discrete random variable**, $g(X)$, defines another discrete random variable whose PMF can be found from that of X.

Probability Mass Functions (PMF)

If x is a possible value of X, the **probability mass** of x is the probability of the event $\{X = x\}$ consisting of all outcomes that result in X taking the value x:

$$p_X(x) = \mathbf{P}(\{X = x\}).$$

Note,

1. $\sum_x p_X(x) = 1$, follows from the additivity and normalization axioms of probability.
2. $\mathbf{P}(X \in S) = \sum_{x \in S} p_X(x)$, for any set S of possible values of X.

Calculation of Probability Mass Functions

Calculating the PMF of X :

1. Collect all possible outcomes in Ω which are mapped by X to x, i.e. those that give rise to the event $\{X = x\}$.
2. Add their probabilities to obtain $p_X(x)$.

The PMF of a function of a random variable, $Y = g(X)$, can be found from the PMF of X:

$$p_Y(y) = \sum_{\{x|g(x)=y\}} p_X(x).$$

Expected Value and Variance

Expected Values

- of X whose PMF is $p_X(x)$: $\mathbf{E}[X] = \sum_x x p_X(x)$.
- of $g(X)$ can be found using $p_X(x)$, and is $\mathbf{E}[g(X)] = \sum_x g(x) p_X(x)$.
- The n**th moment** of X is $\mathbf{E}[X^n] = \sum_x x^n p_X(x)$.
- If $Y = aX + b$, then $\mathbf{E}[Y] = a\mathbf{E}[X] + b$.

Variance of X whose PMF is $p_X(x)$:

- $\text{var}(X) = \mathbf{E}[(X - \mathbf{E}[X])^2] = \sum_x (x - \mathbf{E}[X])^2 p_X(x)$.
- $\text{var}(X) \geq 0$, and $\sqrt{\text{var}(X)} = \sigma_X$ is the **standard deviation**.
- $\text{var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$.
- If $Y = aX + b$, then $\text{var}(Y) = a^2 \text{var}(X)$.

Some Discrete Random Variables

	X	$p_X(k)$	$\mathbf{E}[X]$	$\text{var}(X)$
Bernoulli	$\begin{cases} 1, & \text{if H,} \\ 0, & \text{if T.} \end{cases}$ (Coin toss)	$\begin{cases} p, & k=1, \\ 1-p, & k=0. \end{cases}$	p	$p(1-p)$
Binomial	# of H in n tosses.	$\binom{n}{k}p^k(1-p)^{n-k}$ $k=0,1,\ldots,n.$	np	$np(1-p)$
Geometric	# of tosses til 1st head.	$(1-p)^{k-1}p$ $k=1,2,\ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson		$\frac{\lambda^k}{k!}e^{-\lambda}$ $k=1,2,\ldots$	λ	λ
Uniform	$p_X(k) = \begin{cases} \frac{1}{b-a+1}, & k=a,\ldots b, \\ 0, & \text{otherwise.} \end{cases}$		$\frac{a+b}{2}$	$\frac{(b-a)(b-a+2)}{12}$

Multiple Random Variables

Let X and Y be random variables associated with the same experiment.

- The **joint PMF** of X and Y is $p_{X,Y}(x,y) = \mathbf{P}(X = x, Y = y)$.
- The **marginal PMF**s of X and Y can be obtained from the joint PMF: $p_X(x) = \sum_y p_{X,Y}(x,y)$, and $p_Y(y) = \sum_x p_{X,Y}(x,y)$.
- $Z = g(X,Y)$ defines a new random variable:
  - $p_Z(z) = \sum_{\{(x,y)|g(x,y)=z\}} p_{X,Y}(x,y)$.
  - $\mathbf{E}[g(X,Y)] = \sum_x \sum_y g(x,y) p_{X,Y}(x,y)$.
  - If g is linear, i.e. $g(X,Y) = aX + bY + c$ for a, b, c scalars, then, $\mathbf{E}[g(X,Y)] = a\mathbf{E}[X] + b\mathbf{E}[Y] + c$.

Conditioning

- **Conditioning** X **on an event** A, with $\mathbf{P}(A) > 0$ results in the PMF

$$p_{X|A}(x) = \mathbf{P}(X = x|A) = \frac{\mathbf{P}(\{X = x\} \cap A)}{\mathbf{P}(A)}.$$

  Note, $\sum_x p_{X|A}(x) = 1$; $p_{X|A}(x)$ is a legitimate PMF.

- **Conditioning** X **on** $\{Y = y\}$, with $P_Y(y) > 0$ results in the PMF

$$p_{X|Y}(x|y) = \mathbf{P}(X = x|Y = y) = \frac{\mathbf{P}(X = x, Y = y)}{\mathbf{P}(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}.$$

  Note,

  - $p_{X,Y}(x, y) = p_{X|Y}(x|y)p_Y(y)$, (recall multiplication rule),
  - $p_X(x) = \sum_y p_{X|Y}(x|y)p_Y(y)$, (recall total probability thm).

Conditional Expectation

Let X and Y be random variables associated with the same experiment.

- The **conditional expectation** of X,
  - given an event A, $\mathbf{P}(A) > 0$, is $\mathbf{E}[X|A] = \sum_x x p_{X|A}(x)$.
  - given $Y = y$, $P_Y(y) > 0$, is $\mathbf{E}[X|Y = y] = \sum_x x p_{X|Y}(x|y)$.
- For $g(X)$, $\mathbf{E}[g(X)|A] = \sum_x g(x) p_{X|A}(x)$.
- **Total Expectation Theorem**
  - $\mathbf{E}[X] = \sum_y p_Y(y)\mathbf{E}[X|Y = y]$.
  - Let $A_1, \ldots, A_n$ be disjoint events that partition Ω and $\mathbf{P}(A_i) > 0$, $\forall i$. Then, $\mathbf{E}[X] = \sum_{i=1}^n \mathbf{P}(A_i)\mathbf{E}[X|A_i]$.
  - Let $A_1, \ldots, A_n$ be disjoint events that partition B and $\mathbf{P}(A_i \cap B) > 0$, $\forall i$. $\mathbf{E}[X|B] = \sum_{i=1}^n \mathbf{P}(A_i|B)\mathbf{E}[X|A_i \cap B]$.

Independence

Let X and Y be random variables associated with the same experiment and A an event with $\mathbf{P}(A) > 0$.

- X is independent of A if
  - $p_{X|A}(x) = p_X(x)$, $\forall x$, or,
  - $\mathbf{P}(X = x \text{ and } A) = p_X(x)\mathbf{P}(A)$, $\forall x$.
- X and Y are independent if
  - $p_{X,Y}(x, y) = p_X(x)p_Y(y)$, $\forall x, y$, or,
  - $p_{X|Y}(x|y) = p_X(x)$, $\forall x$ and $\forall y$ such that $p_Y(y) > 0$.
- If X and Y are independent,
  - then, so are $g(X)$ and $h(Y)$.
  - then, $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$. *(The converse is* **not** *true;* $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y] \nRightarrow X$ *and* Y *are independent.)*
  - then, $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$.

Continuous Random Variables: Probability Density Functions (PDF)

For a continuous RV X with PDF $f_X(x)$ (≥ 0),

$$\begin{aligned} P(a \leq X \leq b) &= \int_a^b f_X(x)dx \\ P(x \leq X \leq x+\delta) &\approx f_X(x).\delta \\ P(X \in A) &= \int_A f_X(x)dx \end{aligned}$$

Remarks:

- if X is continuous, $P(X = x) = 0 \ \ \forall x$!!
- $f_X(x)$ may take values larger than 1.

Normalization property:

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

Mean and variance of a continuous RV

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f_X(x) dx \\
E[g(X)] &= \int_{-\infty}^{\infty} g(x) f_X(x) dx \\
\mathrm{Var}(X) &= \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx \\
&= E[X^2] - (E[X])^2 \ (\geq 0) \\
E[aX + b] &= aE[X] + b \\
\mathrm{Var}(aX + b) &= a^2 \mathrm{Var}(X)
\end{aligned}
$$

Cumulative Distribution Functions

Definition:

$$F_X(x) = P(X \leq x)$$

monotonically increasing from 0 (at $-\infty$) to 1 (at $+\infty$).

- Continuous RV:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt \quad \text{(continuous)}$$

$$f_X(x) = \frac{dF_X}{dx}(x)$$

- Discrete RV:

$$F_X(x) = P(X \leq x) = \sum_{k \leq x} p_X(k) \quad \text{(piecewise constant)}$$

$$p_X(k) = F_X(k) - F_X(k-1)$$

Normal/Gaussian Random Variables

Standard Normal RV: $N(0,1)$:

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \\ E[X] &= 0, \qquad \mathrm{Var}(X) = 1 \end{aligned}$$

General normal RV: $N(\mu, \sigma^2)$:

$$\begin{aligned} f_X(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \\ E[X] &= \mu, \qquad \mathrm{Var}(X) = \sigma^2 \end{aligned}$$

- if $Y = aX + b$, then $Y \sim N(a\mu + b,\ a^2\sigma^2)$.
- CDF for standard normal $\phi(.)$ can be read in a table.
- To evaluate CDF of a general standard normal, express it as a function of a standard normal:

$$X \sim N(\mu, \sigma^2) \Leftrightarrow \frac{X - \mu}{\sigma} \sim N(0, 1)$$

$$P(X \leq x) = P\Big(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\Big) = \phi\Big(\frac{x - \mu}{\sigma}\Big)$$

where $\phi(.)$ denotes the CDF of a standard normal.

Some Continuous Random Variables

	$f_X(x)$	$F_X(x)$	$\mathbf{E}[X]$	$\text{var}(X)$
Uniform ([a, b])	$\begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0 & \text{o.w.} \end{cases}$	$\begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & \text{o.w.}(x > b) \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential (λ)	$\begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{o.w.} \end{cases}$	$\begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & \text{o.w.} \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal (μ, σ^2)	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$		μ	σ^2