T2

Probability Theory

Salvador Ruiz Correa

August 18, 2025

This section introduces the foundational concepts of probability distributions and their continuous and discrete representations. We begin by defining probability distributions as mathematical functions that assign likelihoods to outcomes within a sample space. For discrete random variables, this takes the form of a probability mass function (PMF), while for continuous random variables, it is described by a probability density function (PDF). We explore the properties and interpretations of these functions, emphasizing their roles in quantifying uncertainty and modeling real-world phenomena. The discussion then extends to cumulative distribution functions (CDFs), which capture the probability that a random variable takes a value less than or equal to a given threshold. Both discrete and continuous cases are treated, highlighting the differences in their behavior and analytical structure. Key relationships between PMFs, PDFs, and their respective CDFs are established, and graphical illustrations are used to reinforce intuition. By the end of this section, readers will have a clear understanding of how probability distributions and densities encapsulate randomness, and how cumulative functions provide a powerful tool for analyzing probabilistic events over intervals.

Agenda:
1 Probability distribution function.
2 Probability density function.
3 Cumulative dstribution function.
4 Cummulative density function

Examples of Probability Distributions for Discrete Random Variables

- Let $(\Omega, \mathscr{F})$ describe throwing two fair dice, i.e. $\Omega := \{(i,k) : 1 \leq i,k \leq 6\}$, $\mathscr{F} = \mathscr{P}(\Omega)$, and $P(\{i,j\}) = \frac{1}{36}$. The total number of points thrown $X : \Omega \to \{2,3,\ldots,12\}$, $X((i,j)) = i + j$ is a measurable map (Table 1 and Figure 21).

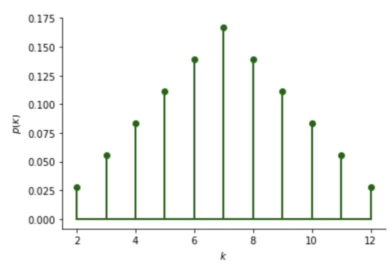

Figure 1: Probability distribution $P(K = k)$ in Table 1.

k	2	3	4	5	6	7	8	9	10	11	12
$P(X = k)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Table 1: The distribution of the random variable X.

- The *Bernoulli random* variable $X \in \{0,1\}$ with parameter p, $0 < \mathsf{p} \leq 1$ has the following probability distribution:

$$P(X = x \mid \mathsf{p}) := \mathsf{Bernoulli}(X = x \mid \mathsf{p}) = \mathsf{p}^x(1-\mathsf{p})^{1-x}.$$

Hint: $P(X = 1 \mid \mathsf{p}) = \mathsf{p}$.

- The *Binomial distribution* with parameters N and p is the discrete probability distribution of the number K of successes in a sequence of N independent Bernoulli trials (with parameter p). The probability distribution is

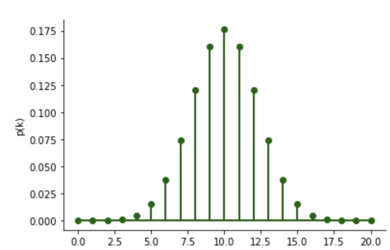

Figure 2: Binomial probability distribution function for $N = 20$ and $p = 0.5$.

$$P(K = k \mid \mathsf{N}, \mathsf{p}) := \mathsf{Binomial}(K = k \mid \mathsf{p}, \mathsf{N}) = \binom{\mathsf{N}}{k} \mathsf{p}^k (1 - \mathsf{p})^{N-k}$$

for $\quad k = 0, 1, 2, \ldots$ where

$$\binom{\mathsf{N}}{k} = \frac{\mathsf{N}!}{(\mathsf{N} - k)!k!}$$

is the number of ways of choosing $K = k$ objects out of a total of N identical objects.

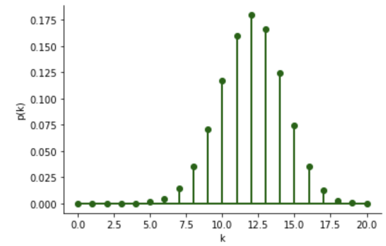

Figure 3: Binomial probability distribution function for $N = 20$ and $p = 0.6$.

Continuous Random Variables

A random variable X is a continuous random variable if there exists a non-negative function $f_X(\cdot)$ such that:

$$P(X \leq \omega) = \int_{-\infty}^{\omega} f_X(\alpha) d\alpha$$

for any $\omega \in \mathbb{R}$. The function f_X is called the probability density function of X. To simplify notation in practice we use $p(x) = f(x) = f_X(x)$. We remark on the abuse of notation.

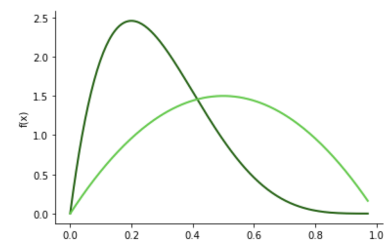

Figure 4: Beta probability density function. $\alpha = 2, \beta = 5$ (dark green), $\alpha = 2, \beta = 2$ (lime green).

Examples of Continuous Random Variables

- A random variable $M \in [0, 1]$ has a Beta distribution of variable with parameters α and β if the density function has the form

$$f_M(m \mid \alpha, \beta) := \mathsf{Beta}(m \mid \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} m^{\alpha - 1}(1 - m)^{\beta - 1},$$

  where $\Gamma(x)$ is the Gamma function $\Gamma(x) = \int_0^x u^{x-1} e^{-u} du$.
- A random variable $X \in \mathbb{R}$ has a Gaussian or Normal distribution of variable with parameters μ and σ^2 if the density function has the form

$$f_X(x \mid \mu, \sigma^2) := \mathcal{N}(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x - \mu)^2\},$$

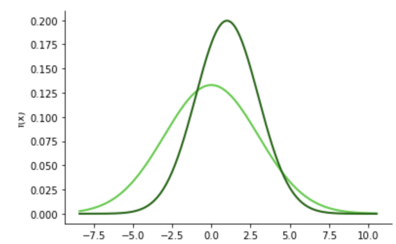

Figure 5: Gaussian probability density function. $\mu = 0, \sigma^2 = 3$ (dark green), $\mu = 2, \sigma^2 = 2$ (lime green).

Cummulative Distribution Function

For a random variable X, its cumulative distribution function (CDF) is defined as:

$$F_X(x) := P(X \leq x), \quad -\infty \leq x \leq \infty$$

Note that $P(X \leq x) = P \circ X^{-1}((-\infty, x])$, and:
- $F_X(x)$ is non-decreasing and right-continuous.

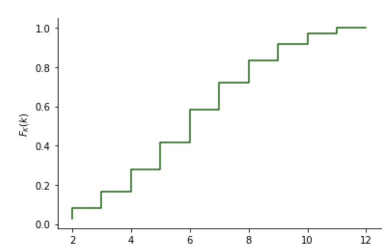

Figure 6: Cumulative distribution of the random variable X in Table 1.

- $\lim_{x\to\infty} F_X(x) = 1$ and
- $\lim_{x\to-\infty} F_X(x) = 0$

Conversely, if a given function F_X satisfies the above properties, then it is a CDF of some random variable.

As an example, we show below the cumulative distribution of the random variable X in Table 1 (see Figures 21 and 22).

k	2	3	4	5	6	7	8	9	10	11	12
$F_K(k)$	0.02	0.08	0.16	0.27	0.41	0.58	0.72	0.83	0.91	0.97	1

Box 1: Discrete and continuous random variables

- **Bernoulli** ($X \sim \text{Bernoulli}(p)$) $\mathbb{P}(X = x) = p^x(1-p)^{1-x}, \quad x \in \{0,1\}$
- **Binomial** ($X \sim \text{Binomial}(n,p)$) $\mathbb{P}(X = k) = \binom{n}{k}p^k(1-p)^{n-k}, \quad k = 0,1,\ldots,n$
- **Geometric** ($X \sim \text{Geometric}(p)$) $\mathbb{P}(X = k) = (1-p)^{k-1}p, \quad k = 1,2,\ldots$
- **Negative Binomial** ($X \sim \text{NegBin}(r,p)$) $\mathbb{P}(X = k) = \binom{k-1}{r-1}p^r(1-p)^{k-r}, \quad k = r, r+1, \ldots$
- **Poisson** ($X \sim \text{Poisson}(\lambda)$) $\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0,1,2,\ldots$
- **Hypergeometric** ($X \sim \text{Hypergeo}(N,K,n)$) $\mathbb{P}(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad k = \max(0, n-N+K), \ldots, \min(n,K)$
- **Discrete Uniform** ($X \sim \text{Unif}\{a, a+1, \ldots, b\}$) $\mathbb{P}(X = x) = \frac{1}{b-a+1}, \quad x = a, a+1, \ldots, b$
- **Uniform** ($X \sim \text{Uniform}(a,b)$) $f(x) = \frac{1}{b-a}, \quad x \in [a,b]$
- **Exponential** ($X \sim \text{Exponential}(\lambda)$) $f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$
- **Normal (Gaussian)** ($X \sim \mathcal{N}(\mu, \sigma^2)$) $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
- **Gamma** ($X \sim \text{Gamma}(\alpha,\beta)$) $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x}, \quad x > 0$
- **Beta** ($X \sim \text{Beta}(\alpha,\beta)$) $f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \quad x \in (0,1)$
- **Chi-Squared** ($X \sim \chi^2(k)$) $f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1}e^{-x/2}, \quad x > 0$
- **Student's t** ($X \sim t_\nu$) $f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$
- **Cauchy** ($X \sim \text{Cauchy}(x_0, \gamma)$) $f(x) = \frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]}$

Box 2: Multinomial Random Variable

Let $X = (X_1, X_2, \dots, X_k)$ be a random vector representing counts of outcomes in k categories.

- **Definition:** $X \sim \text{Multinomial}(n; p_1, p_2, \dots, p_k)$ where n is the number of trials, and p_i is the probability of category i, with $\sum_{i=1}^{k} p_i = 1$.
- **Probability Mass Function (PMF):**

$$\mathbb{P}(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

Derivation: Poisson as Limit of Binomial

Start with the Binomial probability mass function:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Let $\lambda = np$, and consider the limit as $n \to \infty$, $p \to 0$, with λ fixed. Rewrite the PMF in terms of λ:

$$P(X = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Approximate the binomial coefficient for fixed k and large n:

$$\binom{n}{k} \approx \frac{n^k}{k!}$$

Now substitute and simplify:

$$P(X = k) \approx \frac{n^k}{k!} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k}$$

Note the limits:

$$\left(\frac{\lambda}{n}\right)^k = \frac{\lambda^k}{n^k}, \quad \left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}, \quad \left(1 - \frac{\lambda}{n}\right)^{-k} \to 1$$

Putting it all together:

$$P(X = k) \approx \frac{n^k}{k!} \cdot \frac{\lambda^k}{n^k} \cdot e^{-\lambda} = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$
\begin{aligned}
P(X=k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^{k}\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n(n-1)(n-2)\dots(n-k+1)(n-k)!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^{k}\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!}\left(\frac{n(n-1)(n-2)\dots(n-k+1)}{n^k}\right)\left(1-\frac{\lambda}{n}\right)^{-k}\left(1-\frac{\lambda}{n}\right)^{n} \\
&= \frac{\lambda^k}{k!}\left(\frac{n(n-1)(n-2)\dots(n-k+1)}{n\cdot n\cdot n\cdots n}\right)\left(1-\frac{\lambda}{n}\right)^{-k}\left(1-\frac{\lambda}{n}\right)^{n}
\end{aligned}
$$

$$
\begin{aligned}
\lim_{n\to\infty} P(X=k) &= \lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda}{n}\right)^{k}\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty}\frac{n(n-1)(n-2)\dots(n-k+1)(n-k)!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^{k}\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty}\frac{\lambda^k}{k!}\left(\frac{n(n-1)(n-2)\dots(n-k+1)}{n^k}\right)\left(1-\frac{\lambda}{n}\right)^{-k}\left(1-\frac{\lambda}{n}\right)^{n} \\
&= \frac{\lambda^k}{k!}\lim_{n\to\infty}\left(\frac{n(n-1)(n-2)\dots(n-k+1)}{n\cdot n\cdot n\cdots n}\right)\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} \\
&= \frac{\lambda^k}{k!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} \\
&= \frac{\lambda^k}{k!}\lim_{n\to\infty}\left(1+\frac{1}{a}\right)^{-a\lambda} \\
&= \frac{\lambda^k}{k!}e^{-\lambda}
\end{aligned}
$$

with $a = -\lambda/n$.

Thus, the Binomial distribution converges to the Poisson distribution:

$$\boxed{P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}}$$

Drawing Samples from Random Variable

Drawing samples from a random variable refers to the process of generating individual realizations or instances of that random variable according to its probability distribution. In probability theory and statistics, a random variable is a variable whose possible values are outcomes of a random phenomenon. The probability distribution of a random variable describes the likelihood of different values it can take.

When you draw samples from a random variable, you are essentially simulating or generating data points that follow the probability distribution of that variable. This process is often used in various fields such as statistics, machine learning, and simulations to understand the behavior of random phenomena or to make predictions.

For example, if you have a random variable representing the outcome of a fair six-sided die roll, drawing samples from this random variable would involve simulating the roll of the die and obtaining values like 1, 2, 3, 4, 5, or 6 with equal probability.

The concept of drawing samples is fundamental to Monte Carlo simulations, where random sampling is used to estimate numerical results and analyze complex systems that involve randomness. In statistical terms, the more samples you draw, the better your approximation of the true underlying distribution or properties of the random variable.