Probability Theory and Statistics

Lecture 9

Bence Csonka

Budapest University of Technology and Economics
csonkab@edu.bme.hu

October 14

Remainder Continuous Random Variables

Definition

Let $X : \Omega \to \mathbb{R}$ *be a continuous random variable. The distribution function of* X *is defined by*

$$F_X(x) = \mathbb{P}(X < x).$$

The density function of X *is the function which satisfies*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, \mathrm{d}t.$$

Remainder: Expectation in the Continuous Case

The continuous analogue: instead of summing the weight function times k, we integrate the density times x—this is also a "probability-weighted average".

Definition

Let X be a continuous random variable such that

$$\int_{-\infty}^{\infty} |x| f_X(x)\mathrm{d}x < \infty.$$

Then

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f_X(x)\mathrm{d}x.$$

Expectation of a Transform

In the continuous case, quantities like the variance of X are important:

$$\mathbb{D}^2(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Question: How do we compute $\mathbb{E}(X^2)$?

Expectation of a Transform

In the continuous case, quantities like the variance of X are important:

$$\mathbb{D}^2(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Question: How do we compute $\mathbb{E}(X^2)$?

1. Option: apply the following statement with $g(x) = x^2$.

Theorem

Let X be a continuous random variable and $g\colon \mathbb{R} \to \mathbb{R}$ a continuous function. If

$$\int_{-\infty}^{\infty} |g(x)|\, f_X(x)\, \mathrm{d}x < \infty,$$

then

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, \mathrm{d}x.$$

Remark: If $g(x) \geq 0$ for all $x \in \mathrm{Ran}(X)$, then $\mathbb{E}(g(X)) \in [0, \infty]$ is always well-defined. For instance, $\mathbb{E}(X^2)$ is always well-defined (possibly infinite).

Distribution of a Transform in the Continuous Case

Question: How do we compute $\mathbb{E}(X^2)$?

2. Option: Determine the density $f_{g(X)}$ of the random variable $g(X)$, then compute its expectation by the usual formula:

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} x \cdot f_{g(X)}(x)\mathrm{d}x.$$

If we only need $\mathbb{E}(g(X))$, it is often easier to use the formula on the previous slide. But if we also want the distribution of $g(X)$, we must find $f_{g(X)}$. Procedure:

1. Determine the range $\mathrm{Ran}(g(X))$.
2. Compute the distribution function $F_{g(X)}$ by definition.
3. Differentiate it (where appropriate) to obtain the density $f_{g(X)}$.

Caution: $f_{g(X)}$ typically cannot be obtained directly.

Transform of a Uniform (the Square)

Example: the distribution of $Y = X^2$ where X is uniform on $(0,1)$.

1. $\mathrm{Ran}(Y) = (0,1)$, since $\{x^2 \colon x \in (0,1)\} = (0,1)$.
2. $F_Y(x) = \begin{cases} 0, & \text{if } x \le 0, \\ \sqrt{x}, & \text{if } 0 < x < 1, \\ 1, & \text{if } x \ge 1. \end{cases}$

   (Computation: see lecture.)
3. Hence $f_Y(x) = \begin{cases} \dfrac{1}{2\sqrt{x}}, & \text{if } 0 < x < 1, \\ 0, & \text{if } x \le 0 \text{ or } x \ge 1. \end{cases}$

Properties of the Variance

These remain true in the continuous/general case: if $c \in \mathbb{R}$ and $\mathbb{E}(X^2) < \infty$, then $\mathbb{D}^2(cX) = c^2\mathbb{D}^2(X)$ and $\mathbb{D}^2(X + c) = \mathbb{D}^2(X)$.

Reminder: (6.4.2. Proposition) If X and Y are independent random variables with $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$, then

$$\mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y).$$

We proved this only in the discrete case (and only partially), but the proof is analogous for continuous variables.

The key point in the discrete proof: if X, Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. This holds for *any* independent random variables.

- We already know what independence means for general random variables: X, Y are independent $\Leftrightarrow$ $\mathbb{P}(X < x, Y < y) = \mathbb{P}(X < x)\mathbb{P}(Y < y)$, for all $x, y \in \mathbb{R}$.
- We have not yet discussed how to compute $\mathbb{E}(XY)$ in the continuous case (nor precisely what "continuous case" means for joint distributions) $\rightarrow$ coming soon.

Variances of Notable Continuous Distributions

Theorem

If X is uniformly distributed on (a, b), then

$$\mathbb{D}^2(X) = \frac{(b-a)^2}{12}.$$

(Thus $\mathbb{D}(X) = \frac{b-a}{2\sqrt{3}}$.)

(Proof — compute $\mathbb{E}(X^2)$: see lecture.)

Variance of Exponential Distribution

Theorem

If $Z \sim \mathrm{Exp}(\lambda)$ *(with* $\lambda > 0$*), then* $\mathbb{D}^2(Z) = \frac{1}{\lambda^2}$ *(hence* $\mathbb{D}(Z) = \frac{1}{\lambda}$*).*

(Proof: we already know that $\mathbb{E}(Z)^2 = \frac{1}{\lambda^2}$, so it suffices to show $\mathbb{E}(Z^2) = \frac{2}{\lambda^2}$.
The computation is analogous to that of $\mathbb{E}(Z)$: perform integration by parts twice with the same roles; or avoid the second integration by recognizing the $\mathrm{Exp}(\lambda)$ density.

Simulation of Continuous Random Variables

An application of distribution transformations in computer science:

Assume that we have a random number generator capable of producing approximately independent, approximately $\mathrm{U}(0;1)$ distributed (pseudo)random numbers.

The generation of pseudorandom numbers is beyond the scope of this course; we can treat it as a black box.

Question: How can we use this to simulate (approximately) independent random variables following (approximately) a given continuous distribution? For example, $\mathrm{Exp}(\lambda)$ distributed ones?

Simulation of Continuous Random Variables

An application of distribution transformations in computer science:

Assume that we have a random number generator capable of producing approximately independent, approximately $\mathrm{U}(0;1)$ distributed (pseudo)random numbers.

The generation of pseudorandom numbers is beyond the scope of this course; we can treat it as a black box.

Question: How can we use this to simulate (approximately) independent random variables following (approximately) a given continuous distribution? For example, $\mathrm{Exp}(\lambda)$ distributed ones?

Answer: Substitute the $\mathrm{U}(0;1)$ random variables into the inverse of the target distribution function.

We will justify the correctness of this procedure in two steps.

Simulation of Continuous Random Variables

Theorem

If X is a continuous random variable, then $F_X(X)$ is $\mathrm{U}(0;1)$ distributed.

Simulation of continous random variable

Theorem

If X is a continuous random variable and $U \sim \mathrm{U}(0;1)$, then the distribution function of $F_X^{-1}(U)$ is F_X.

(Proof in the above case)
($F_X^{-1}(U)$ has distribution function $F_X \Leftrightarrow F_X^{-1}(U)$ and X have the same distribution.)

Remark: The Statement and the Theorem remain true for any continuous random variable, but more care is needed to specify on which domain F_X and F_X^{-1} are strictly monotone.

Simulation of Continuous Random Variables: Example

If $X \sim \mathrm{Exp}(\lambda)$ with $\lambda > 0$, then $F_X(x) = 1 - \mathrm{e}^{-\lambda x}$.
The density of X is positive on the open interval $I = (0, \infty)$ and zero elsewhere.
$\lim_{x \downarrow 0} F_X(x) = 0$, $\lim_{x \uparrow \infty} F_X(x) = 1$, and F_X is strictly increasing on $(0, \infty)$.
Thus for $x > 0$ and $u \in (0, 1)$,

$$
\begin{aligned}
x = {F_X}^{-1}(u) \Leftrightarrow F_X(x) = u \qquad & \Leftrightarrow 1 - \mathrm{e}^{-\lambda x} = u \\
\Leftrightarrow -\lambda x = \ln(1 - u) \quad & \Leftrightarrow x = \frac{-\ln(1-u)}{\lambda}.
\end{aligned}
$$

Therefore, if $U \sim \mathrm{U}(0; 1)$ is a random number generated by a computer, then $\frac{-\ln(1-U)}{\lambda} \sim \mathrm{Exp}(\lambda)$.

Simulation of Discrete Random Variables

For discrete random variables X, the previous method cannot be used, since F_X is not invertible (it is piecewise constant with jumps in between).
Question: How can we simulate X using a random variable $U \sim \mathrm{U}(0;1)$?

Simulation of Discrete Random Variables

For discrete random variables X, the previous method cannot be used, since F_X is not invertible (it is piecewise constant with jumps in between).
Question: How can we simulate X using a random variable $U \sim \mathrm{U}(0;1)$?

Answer: If $\mathrm{Ran}(X) = \{k_1, k_2, \ldots\}$ (where the k_i are distinct), divide the interval $(0,1)$ into disjoint subintervals $I_1, I_2, \ldots$ such that the length of I_i equals $\mathbb{P}(X = k_i)$.
Then we simulate X as follows: if $U \in I_i$, we return the value $X = k_i$.

Example: for a fair die, $U \in (0, 1/6] \Rightarrow X = 1$, $U \in (1/6, 2/6] \Rightarrow X = 2$, $\ldots$, $U \in (5/6, 1) \Rightarrow X = 6$. This is a perfect simulation (at least as perfect as the $\mathrm{U}(0;1)$ generator itself).

If the range of X is infinite (e.g. geometric, Poisson, or Zipf distributions), only finitely many possible values can be simulated in this way.

End of the midterm material.

- From the exercise sheets, the 6th sheet is the last one included in the midterm (including everything from it even if it is discussed during the 6th week).
- "Theoretical questions" (definitions, theorem statements) will only appear in the final exam, but of course you will need to use the definitions and results from the lectures in the midterm tasks. You must refer to and name the relevant theorems and properties.
  For example: "The events A, B and C form a complete system of events, since they are pairwise disjoint and their union is Ω (with justification). Therefore, by the Law of Total Probability: ..."
  Naturally, shorter phrasing is acceptable, but the concept of a complete system of events and the LTP must appear.
- You may bring and use a calculator (recommended). We will provide tables for the standard distributions; you may not use your own.