

Continuous Random Variables

Math 394

1 (Almost bullet-proof) Definition of Expectation

Assume we have a sample space Ω , with a σ -algebra of subsets \mathcal{F} , and a probability P , satisfying our axioms. Define a random variable as a function $X : \Omega \rightarrow \mathbb{R}$, such that all subsets of Ω of the form $\{\omega \mid a < X(\omega) \leq b\}$, for any real $a \leq b$ are events (belong to \mathcal{F}). Assume at first that the range of X is bounded, say it is contained in the interval $[A, B]$. We work with X by approximating it with a sequence of discrete random variables $X^{(n)}$, defined as

$$\{X^{(n)} = x_k\} = \{x_k \leq X < x_{k+1}\}$$

where $A = x_0 < x_1 < \dots < x_n = B$ is a partition of our interval with, for example, $|x_{j+1} - x_j| = \frac{B-A}{n}$.

We can now define

$$E[X] = \lim_{n \rightarrow \infty} E[X^{(n)}] = \lim_{n \rightarrow \infty} \sum_{k=1}^n x_k P[x_k \leq X < x_{k+1}]$$

if the limit exists.

We limit ourselves to absolute continuous random variables, so that

$$P[x_k \leq X < x_{k+1}] = \int_{x_k}^{x_{k+1}} f_X(u) du$$

where f_X is a piecewise continuous non-negative function, such that $\int_{-\infty}^{\infty} f_X(x) dx = \int_A^B f_X(x) dx = 1$. It is now straightforward to prove that $E[X] = \int_{-\infty}^{\infty} x f(x) dx$.
Indeed

$$\begin{aligned} \left| \sum_k x_k \int_{x_k}^{x_{k+1}} f(u) du - \int_A^B x f(x) dx \right| &\leq \sum_k \left| \int_{x_k}^{x_{k+1}} (x_k - x) f(x) dx \right| \leq \\ &\leq \sum_k \int_{x_k}^{x_{k+1}} |x_k - x| f(x) dx \leq \frac{B-A}{n} \int_A^B f(x) dx = \frac{B-A}{n} \rightarrow 0 \end{aligned}$$

repeatedly using the triangle inequality.

If the range of X is unbounded, we proceed as in the definition of improper integrals over the real line, by considering an increasing sequence of intervals $[A_n, B_n]$, with $\bigcup_{n=1}^{\infty} [A_n, B_n] = \mathbb{R}$, and define

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \lim_{n \rightarrow \infty} \int_{A_n}^{B_n} xf(x)dx \quad (1)$$

if the limit exists in the sense of improper integrals (for example by computing separately $\lim_{n \rightarrow \infty} \int_0^{B_n} xf(x)dx$, and $\lim_{n \rightarrow \infty} \int_{A_n}^0 xf(x)dx$ and defining the sum as (1), as long they don't diverge both).

Similarly, we will define $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)dx$ if the limit exists, using the same argument.

Note that most of the results about expectations that we saw in previous chapters extend to the continuous case, thanks to its nature as a limit. Thus, for example,

$$E \left[\sum_{k=1}^n a_k X_k + b \right] = a_k \sum_{k=1}^n E[X_k] + b$$

$$Var \left[\sum_{k=1}^n a_k X_k + b \right] = \sum_{k=1}^n a_k^2 Var[X_k] + 2 \sum_{1=k < j}^n a_k a_j Cov[X_k X_j]$$

with moments, variance, covariance, and so on defined as previously. Also, the proof of Markov's Inequality, and Chebyshev's Inequality go through, as does the proof of the Weak Law of Large Numbers.

Remark Note that we have as a consequence that $E[X - Y] = E[X] - E[Y]$, but $Var[X - Y] = Var[X] + Var[Y] - 2Cov[X_i X_j]$. In other words, we can reduce the variance of a sum only by choosing the covariance with care. A famous application of this principle is in financial mathematics, where the variance of prices is interpreted as *volatility*, and a measure of risk. The suggestion is then to look for investments that have *negative correlation*, in order to reduce their combined risk.

2 Notable Continuous Distributions as Limits of Notable Discrete Ones

2.1 Uniform Distribution

Take an interval, for example $[0, 1]$. Consider the sequence of discrete uniform random variables with distributions

$$P \left[X^{(n)} = \frac{k}{n} \right] = \frac{1}{n}, \quad k = 1, 2, \dots, n$$

Of course,

$$E \left[X^{(n)} \right] = \frac{1}{n} \sum_{k=1}^n \frac{k}{n} = \frac{n+1}{2n}$$

and

$$E \left[\left(X^{(n)} \right)^2 \right] = \frac{1}{n} \sum_{k=1}^n \frac{k^2}{n^2} = \frac{1}{n^3} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6n^2}$$

$$\text{Var} \left[X^{(n)} \right] = \frac{(n+1)(2n+1)}{6n^2} - \frac{(n+1)^2}{4n^2} = \frac{2(n+1)(2n+1) - 3(n+1)^2}{12n^2} =$$

$$\frac{(n+1)(n-1)}{12n^2} = \frac{n^2-1}{12n^2}$$

The limit of this distribution as $n \rightarrow \infty$ has clearly continuous density of the form

$$u(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

and such a random variable has

$$E[X] = \int_0^1 x dx = \frac{1}{2}$$

$$E[X^2] = \int_0^1 x^2 dx = \frac{1}{3}$$

$$\text{Var}[X] = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

which are, as expected, the limits of the corresponding quantities for $X^{(n)}$.

2.2 From the Geometric to the Exponential Distribution

We saw how to obtain a Poisson distribution as a limit of binomials (the “Law of Rare Events”). In a sequence of Bernoulli trials, consider the time of first “success”:

$$\begin{aligned} P[S = n] &= p(1-p)^{n-1} \\ E[S] &= \sum_{n=1}^{\infty} np(1-p)^{n-1} = \frac{1}{p} \\ Var[S] &= \frac{1-p}{p^2} \end{aligned}$$

as we saw when introducing this distribution. Also,

$$P[S > n] = \sum_{k=n+1}^{\infty} p(1-p)^{k-1} = p(1-p)^n \sum_{k=0}^{\infty} (1-p)^k = (1-p)^n$$

As in the Law of Rare Events, take a time span $[0, t]$, divide the time axis in intervals of length $\frac{t}{n}$, and consider a sequence of Bernoulli trials with probability of success $p_n = \frac{\lambda t}{n}$. As we know, the limit of the number of wins will have a Poisson distribution, with parameter λt . We can determine the probability distribution of the first success (or “arrival”) T from

$$P[T > t] = P[N_t = 0] = e^{-\lambda t} \frac{(\lambda t)^0}{0!} = e^{-\lambda t}$$

This is the survival function of an exponential distribution, as can be seen immediately.

Consistent with this, we can show that the limit of the discrete analog, the geometric distribution, tends to the exponential one. Indeed, we would have that

$$P[S_n > t] = \left(1 - \frac{\lambda}{n}\right)^{nt} \rightarrow e^{-\lambda t}$$

We can prove the same fact by invoking a much more powerful theorem, that we mentioned when discussing the Moment Generating Function and its siblings. The theorem says that if a sequence of moment generating functions converges to a moment generating function, the corresponding distributions converge as well (for example, in the sense that the cumulative distribution functions converge¹)

Indeed, for the geometric distribution,

$$M_G(w) = E[e^{wG}] = \sum_{k=1}^{\infty} e^{wk} p (1-p)^{k-1} = e^w p \sum_{k=1}^{\infty} (e^w (1-p))^{k-1} = \frac{e^w p}{1 - e^w (1-p)} = \frac{e^w}{1 - e^w + p e^w}$$

We have then, taking $p_n = \frac{\lambda}{n}$, that

$$M_{G_n}\left(\frac{w}{n}\right) = \frac{e^{\frac{w}{n}} \frac{\lambda}{n}}{1 - e^{\frac{w}{n}} + \frac{\lambda}{n} e^{\frac{w}{n}}}$$

and since $1 - e^{\frac{w}{n}} = -\frac{w}{n} + o\left(\frac{1}{n}\right)$, and $e^{\frac{w}{n}} \rightarrow 1$,

$$M_{G_n}\left(\frac{w}{n}\right) \rightarrow \frac{\lambda}{\lambda - w} = M(w)$$

Computing the standard quantities for the exponential distribution, we have (with repeated integration by parts, based on $\int x e^{-x} dx = -x e^{-x} + \int e^{-x} dx + C$)

$$E[X] = \lambda \int_0^{\infty} x e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^{\infty} (\lambda x) e^{-\lambda x} d(\lambda x) =$$

$$E[X^2] = \lambda \int_0^{\infty} x^2 e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

$$Var[X] = \frac{1}{\lambda^2}$$

$$M_X(w) = \lambda \int_0^{\infty} e^{wx} e^{-\lambda x} dx = \frac{\lambda}{\lambda - w}$$

(note that the integral converges only for $w < \lambda$), which are the limits of the corresponding quantities for geometric distributions, in the setting of the Law of Rare Events.

¹ To be precise, $F_n(x) \rightarrow F(x)$ at every point x where F is continuous. In practically all our examples, F will be continuous everywhere, so the caveat is not relevant there.

3 From the Binomial (and many others) to the Normal Distribution

De Moivre proved by brute calculation that a sequence of binomial distributions with parameters (n, p) (note that p does not change), as $n \rightarrow \infty$, looks more and more like a Gaussian (Normal) distribution. In the more modern form, as stated by Laplace

$$P \left[a \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq b \right] \approx \Phi(b) - \Phi(a)$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du$ is the cumulative distribution function of the *standard normal distribution*, which cannot be expressed in terms of “elementary” functions.

The original proof is based on taking explicitly the limit of the binomial distribution, and applying *Stirling’s Approximation*

$$\frac{n!}{n^{n+\frac{1}{2}} e^{-n} \sqrt{2\pi}} \rightarrow 1$$

as $n \rightarrow \infty$. De Moivre’s result, in somewhat modernized notation is that, for sufficiently large n

$$P[X_n = k] \approx \frac{1}{\sqrt{2\pi np(1-p)}} \exp \left\{ -\frac{(k - np)^2}{2np(1-p)} \right\}$$

A binomial random variable can be thought as a sum of independent Bernoulli random variables, and, in fact, the convergence to a normal distribution holds in a much more general setting, as was soon recognized by Laplace. The *Central Limit Theorem* states that, for a sequence of independent, identically distributed, random variables X_n , with finite variance σ^2 , and mean μ

$$P \left[\sqrt{n} \frac{\frac{1}{n} \sum_{k=1}^n X_k - \mu}{\sqrt{\sigma^2}} \leq x \right] \rightarrow \Phi(x)$$

This can be read in a number of ways, one of which allows a sharper estimate for the discrepancy between $\frac{1}{n} \sum_{k=1}^n X_k$ and its expectation μ , than what provided by Chebyshev’s Inequality (which, however, has broader applicability). This can also be stated in this form: as $n \rightarrow \infty$, adding up n independent identically distributed random variables, with mean μ and variance σ^2 , if their size is scaled by n (the number of terms in the sum), $\sum_{k=1}^n \frac{X_k}{n} \approx \mu$, a constant (this is the verbal statement of the Law of Large Numbers). Equivalently, $\frac{1}{n} \sum_{k=1}^n (X_k - \mu) \approx 0$, that is adding many small mean zero effects results in cancellation. If, however, we scale by \sqrt{n} their difference from μ , $\sum_{k=1}^n \frac{X_k - \mu}{\sqrt{n}} \approx Z$, where Z is a normal random variable with mean 0 and variance σ^2 . A consequence is that adding many small (but not too small) independent mean zero effects results in a normal distribution, which is the basis for the usual theory of random measurement errors, as well for the Maxwell model of molecular kinetics. Another famous application is in the suggestion to *diversify investment portfolios*, by combining many independent stocks, in relatively small quantities: by Chebyshev’s inequality this should reduce their combined variance, aka *volatility*.

While independence can be slightly relaxed (there is a host of results generalizing the CLT), it bears keeping its importance in mind: if the terms in the sum are significantly dependent, the theorem fails, and ignoring this fact can (and does) lead to serious misapplications of this basic result.