Need for description

How do you say anything sensible about a probability measure on $(\mathbb{R}, \mathbb{B})$?

How do you attach one or more numbers to a probability measure that says something sensible about it's behavior?

In particular, how do you say anything about the difference between two probability measures?

Moments of probability measures

Let ν be a probability measure on $(\mathbb{R}, \mathbb{B})$.

Definition: We say that ν has k'th moment for $k \in \mathbb{N}$ if

$$x \mapsto x^k \,, \qquad x \in \mathbb{R}$$

is ν-integrable. In that case

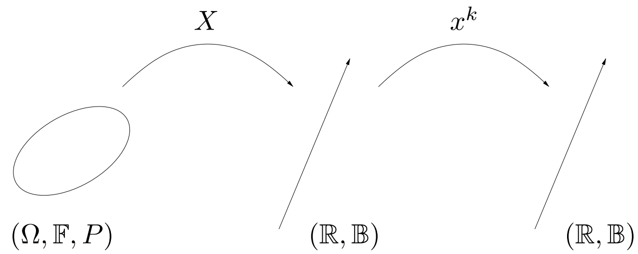
$$\int_{-\infty}^{\infty} x^k \, d\nu(x)$$

is the k'th moment of ν, while

$$\int_{-\infty}^{\infty} |x|^k \, d\nu(x)$$

is the k'th absolute moment of ν.

Moments of stochastic variables



Moments of X are moments of the image measure $\nu = X(P)$ (the distribution). The abstract change-of-variable-formula is

$$\int x^k \, dX(P)(x) = \int X(\omega)^k \, dP(\omega) = \int X^k \, dP$$

For stochastic variables we talk about the expectation of X or in general X^k and write

$$E\,X = \int X \, dP \qquad E\,X^k = \int X^k \, dP$$

Moments of bounded variables

Lemma: If X is a real valued stochastic variable defined on $(\Omega, \mathbb{F}, P)$. If X is almost surely bounded, which means that

$$P(|X| \le C) = 1$$

for some C, then X has moments of all orders.

Proof: Take $k \in \mathbb{N}$, then

$$\int |X|^k \, dP = \int_{(|X| \le C)} |X|^k \, dP \le \int_{(|X| \le C)} C^k \, dP$$

$$= C^k \, P\big(|X| \le C\big)$$

$$< \infty$$

Moments of a continuous distribution

Let X be a real valued random variable defined on $(\Omega, \mathbb{F}, P)$.

Let $t : (\mathbb{R}, \mathbb{B}) \to (\mathbb{R}, \mathbb{B})$ be a measurable transformation.

If the distribution of X has density f w.r.t. m then $t(X)$ has finite expectation (first moment) if and only if

$$\int_{-\infty}^{\infty} |t(x)|\, f(x)\, dx < \infty\,.$$

In which case

$$E\big(t(X)\big) = \int_{-\infty}^{\infty} t(x)\, f(x)\, dx.$$

Empirical moments

Let X be a real valued random variable defined on $(\Omega, \mathbb{F}, P)$.

Let $t : (\mathbb{R}, \mathbb{B}) \to (\mathbb{R}, \mathbb{B})$ be a measurable transformation.

Let $x_1, \ldots, x_n \in \mathbb{R}$ and let X have the empirical distribution, that is

$$X(P) = \epsilon_{x_1,\ldots,x_n}$$

or

$$P(X \in A) = \frac{1}{n}\sum_{i=1}^{n} 1_A(x_i) \qquad \text{for all } A \in \mathbb{B}$$

then $t(X)$ has finite expectation and

$$E\big(t(X)\big) = \frac{1}{n}\sum_{i=1}^{n} t(x_i)$$

Linearity

Lemma: Let X be a real valued stochastic variable defined on $(\Omega, \mathbb{F}, P)$.
If X has first moment, then $\alpha + \beta X$ has first moment for $\alpha, \beta \in \mathbb{R}$, and

$$E\left(\alpha + \beta X\right) = \alpha + \beta E X.$$

Proof: First observe that

$$\int |\alpha + \beta X|\, dP \le \int |\alpha| + |\beta||X|\, dP = |\alpha| + |\beta| \int |X|\, dP < \infty,$$

so $\alpha + \beta X$ has first moment. And

$$E\left(\alpha + \beta X\right) = \int \alpha + \beta X\, dP = \alpha + \beta \int X\, dP = \alpha + \beta\, E X.$$

Large moments imply small moments

Lemma: Let X be a real valued random variable defined on $(\Omega, \mathbb{F}, P)$.
If X has k'th moment then X m'th moment for all $m = 1, \ldots, k$.

Proof: For all $x \in \mathbb{R}$ we have

$$|x|^m \le 1 + |x|^k,$$

therefore

$$|X(\omega)|^m \le 1 + |X(\omega)|^k \quad \text{for all } \omega \in \Omega.$$

This inequality is also written

$$|X|^m \le 1 + |X|^k.$$

Integration on both sides imply

$$\int |X|^m\, dP \le 1 + \int |X|^k\, dP < \infty$$

Additivity

Lemma: Let X and Y be two real valued stochastic variables on $(\Omega, \mathbb{F}, P)$. If X and Y has k'th moment so has $X + Y$.

Proof: For all $x, y \in \mathbb{R}$ we have

$$|x + y| \leq |x| + |y| \leq 2 \max\{|x|, |y|\}.$$

Since $x \mapsto x^k$ is increasing on $[0, \infty)$ we have

$$|x + y|^k \leq 2^k \max\{|x|, |y|\}^k = 2^k \max\{|x|^k, |y|^k\} \leq 2^k(|x|^k + |y|^k).$$

Hence

$$|X + Y|^k \leq 2^k(|X|^k + |Y|^k),$$

and integration yields that

$$\int |X + Y|^k \, dP \leq 2^k \left(\int |X|^k \, dP + \int |Y|^k \, dP \right) < \infty \, .$$

Central moments

If X has k'th moment we introduce that k'th moment around c as

$$E\,(X - c)^k$$

Since $(X - c)^k$ is a polynomium in X of degree k it is integrable.

The k'th central moment of X is the k'th moment around $c = EX$. That is

$$E\,(X - EX)^k.$$

The centrale moments are easier to interpret than the raw moments. The central second moment is called the variance of X and is written $V X$. That is

$$V\,X = E\,(X - E\,X)^2.$$

Formulae

Lemma: Let X be a real valued stochastic variable defined on $(\Omega, \mathbb{F}, P)$. Assume that X has second moment. Then

$$V\,X = E\,X^2 - (E\,X)^2$$

$$V\,X = E\,X^{(2)} - (E\,X)^{(2)}.$$

For all $\alpha, \beta \in \mathbb{R}$ it holds that

$$V\,(\alpha + \beta X) = \beta^2 V X.$$

For all $c \in \mathbb{R}$ it holds that

$$E\,(X - c)^2 = V\,X + (E\,X - c)^2 \, .$$

Interpretation: The expectation, EX, can be characterized as the point around which the distribution of X has the smallest second moment.

Markov's inequality

Theorem: If X is a real valued stochastic variable defined on $(\Omega, \mathbb{F}, P)$. If $P(X \geq 0) = 1$ and X has first moment then for all $c > 0$

$$P(X > c) \leq \frac{EX}{c}.$$

Proof: For all $x \geq 0$ we have

$$c\,1_{(c,\infty)}(x) \leq x.$$

Thereby

$$c\,1_{(c,\infty)}(X) \leq X \qquad P - \text{a.s.}$$

Integration gives that

$$c\,P(X > c) \leq EX.$$

Chebyshev's inequality

Theorem: Let X be a real valued stochastic variable defined on $(\Omega, \mathbb{F}, P)$. If X has second moment then for any $\epsilon > 0$ it holds that

$$P(|X - EX| > \epsilon) \leq \frac{VX}{\epsilon^2} \,.$$

Proof: Define $Y = (X - EX)^2$ and apply Markov's inequality on Y with $c = \epsilon^2$, which yields

$$P(Y > \epsilon^2) \leq \frac{EY}{\epsilon^2}.$$

This is in fact what we want to show since

$$P(Y > \epsilon^2) = P(|X - EX| > \epsilon)$$

and

$$EY = E(X - EX)^2 = VX.$$

Inequalities

Chebyshev's inequality is important because it turns out to be useful with quite minimal assumptions – only the existence of the second moment. Often it does by no means provide a tight bound on the probability $P(|X - EX| > \epsilon)$.

Markov's inequality for positive stochastic variables was underlying and used by applying a transformation.

Let $Y = e^{t(X-EX)}$ for $t \in \mathbb{R}$. If $EY < \infty$ and $c = e^{t\epsilon}$ then

$$P(X - EX > \epsilon) = P(Y > c) \leq e^{-t\epsilon} E e^{t(X-EX)},$$

which for suitable t can become a rather tight inequality – at the expense of integrability of $e^{t(X-EX)}$.

More about the variance

Lemma: Let X be a real valued random variable defined on $(\Omega, \mathbb{F}, P)$. Assume that X has second moment. Then $V\,X \geq 0$ and we have equality if and only if $X = E\,X$ almost surely.

Proof: With $c = EX$ it holds by definition that

$$V\,X = E(X - c)^2 = \int (X - c)^2 \, dP$$

Since $(X - c)^2$ is a non-negative function the integral is ≥ 0 (and $< \infty$ since we assume that X has second moment).

The integral is zero only if the integrand $(X - c)^2$ is P-almost everywhere equal to zero. That is, if $X = c$ almost surely.

□

Standardized variables

When considering higher order moments we often consider the moments of the standardized variables rather than the raw moments.

Define

$$Y = \frac{X - E\,X}{\sqrt{V\,X}}.$$

where $\sqrt{V\,X}$ is called the standard deviation (standardafvigelsen eller spredningen) and is often denoted σ.

Since Y is an affine transformation of X we see that

$$E\,Y = 0\,, \qquad V\,Y = 1.$$

Standardized moments

Let X be a stochastic variable with second moment and let

$$Y = \frac{X - E\,X}{\sqrt{V\,X}}$$

be the standardized variable.

If X has third moment we define the skewness of X, written $\gamma(X)$, as

$$\gamma(X) = E\,Y^3 = \frac{E(X - EX)^3}{(VX)^{3/2}}.$$

If X has fourth moment we define the kurtosis of X, written $\kappa(X)$, as

$$\kappa(X) = E\,Y^4 - 3 = \frac{E(X - EX)^4}{(VX)^2} - 3.$$

The normal distribution $\mathcal{N}(0,1)$ has skewness and kurtosis 0.

Moments in the normal distribution

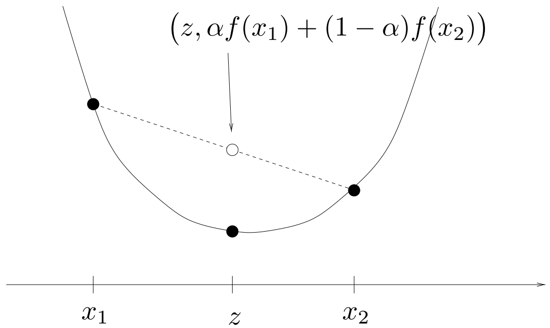
For k even, we can apply the substitution $y = x^2/2$ to compute the k'th moment for the $\mathcal{N}(0,1)$-distribution:

$$\begin{aligned}\int_{-\infty}^{\infty} x^k \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx &= \frac{2^{\frac{k-1}{2}}}{\sqrt{2\pi}}\, 2 \int_0^{\infty} \left(\frac{x^2}{2}\right)^{\frac{k+1}{2}-1} e^{-\frac{x^2}{2}}\,x\,dx \\ &= \frac{2^{\frac{k}{2}}}{\sqrt{\pi}} \int_0^{\infty} y^{\frac{k+1}{2}-1} e^{-y}\,dy \\ &= \frac{2^{\frac{k}{2}}}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right) \\ &= (k-1)(k-3)\cdots 3\cdot 1.\end{aligned}$$

All the even moments are finite, hence so are all the uneven. When k is uneven the integrand is an uneven function, hence the k'th moment is zero.

In particular, if $X \sim \mathcal{N}(\xi, \sigma^2)$ we have $EX = \xi$ and $VX = \sigma^2$.

Convex functions



Definition: A function $f : \mathbb{R} \to \mathbb{R}$ is convex on an interval I if

$$f\big(\alpha x_1 + (1-\alpha)x_2\big) \le \alpha f(x_1) + (1-\alpha) f(x_2)$$

for all $x_1, x_2 \in I$ and $\alpha \in (0,1)$. The function f is called strictly convex on I if the inequality is sharp for all $x_1 \neq x_2$ and all $\alpha \in (0,1)$.

How to check convexity?

Theorem: If $f : \mathbb{R} \to \mathbb{R}$ is differentiable in the open interval I. and if the derivative f' is increasing on I then f is convex on I.

If f' is strictly increasing then f is strictly convex.

Remark: If f is in fact twice differentiable one often checks most easily that f' is increasing or strictly increasing by checking the sign of the second derivative: If

$$f''(x) \ge 0 \quad \text{for all} \quad x \in I$$

then f is convex on I. If

$$f''(x) > 0 \quad \text{for all} \quad x \in I$$

then f is strictly convex on I.

Non-differentiable examples

Theorem: If f_j for $j \in J$ is a family of convex functions on I then if $\sup_{j \in J} f_j$ is finite on I it is convex on I.

Example 1:

$$x \mapsto |x| = \max\{x, -x\}$$

is convex on $\mathbb{R}$.

Example 2: Likewise

$$x \mapsto x^+ = \max\{x, 0\}, \quad x \mapsto x^- = \max\{-x, 0\}$$

are convex functions on $\mathbb{R}$.

Regularity and separation

Theorem: If f is convex on I then f is differentiable from the left and from the right on the interior of I with “increasing derivatives”. Moreover, f is continuous in the interior and measurable (has a measurable extension to $\mathbb{R}$).

Theorem: If $f : \mathbb{R} \to \mathbb{R}$ is a convex function on I, then for every interior point $x \in I$ we can find $\alpha \in \mathbb{R}$ such that

$$f(y) \geq f(x) + \alpha(y - x) \quad \text{for all } \; y \in I\,.$$

If f is strictly convex on I the α can be chosen such that there is sharp inequality for all $y \neq x$.

Jensen’s inequality

Let X be a real valued stochastic variable defined on $(\Omega, \mathbb{F}, P)$ and assume that X has first moment.

Theorem: Let f be a convex function on the interval I, then if $P(X \in I) = 1$ and if X and $f(X)$ have first moment it holds that

$$f(EX) \leq E\big(f(X)\big).$$

If f is strictly convex on I we have sharp inequality unless $X = E\,X$ almost surely.

The proof relies on

Lemma: If $P(X \in I) = 1$ for an interval I and that X does not have a degenerate distribution then $E\,X$ belongs to the interior of I.

Remember Jensen’s inequality

We already know that if X has second moment than

$$VX = EX^2 - (EX)^2 \geq 0$$

which implies that $(EX)^2 \leq EX^2$.

This also follows from Jensen’s inequality applied to the convex function $f(x) = x^2$.

We can always remember Jensen’s inequality because we can never have negative variances!

The inequality tells that the generalized linearity result, which would have been wonderful,

$$f(EX) = E\big(f(X)\big)$$

unfortunately is wrong!

Small and large moments

Theorem: If X has k'th moment then for $m = 1, \ldots, k$

$$\left(E(|X^m|)\right)^{\frac{1}{m}} \leq \left(E(|X^k|)\right)^{\frac{1}{k}}.$$

Proof: Since $\frac{k}{m} > 1$ the function $x \to x^{\frac{k}{m}}$ is convex on $[0, \infty)$. Therefore

$$\left(E(|X|^m)\right)^{\frac{k}{m}} \leq E\left((|X|^m)^{\frac{k}{m}}\right) = E(|X|^k).$$

□

Arithmetic and geometric mean

Let the distribution of X be $\epsilon_{x_1,\ldots,x_n}$ with $x_1, \ldots, x_n > 0$.

Consider the function $f(x) = -\log x$, which is convex on $(0, \infty)$.

Jensen's inequality gives that

$$-\log\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \leq \frac{1}{n}\sum_{i=1}^{n} -\log(x_i) = -\frac{1}{n}\log\left(\prod_{i=1}^{n} x_i\right).$$

Take the exponential on both sides and we get

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i\right)^{1/n},$$

and we have proved that the arithmetic mean of $x_1, \ldots, x_n$ is greater than the geometric mean.