Probabilities and statistics

Lecturer: Prof. Dr. Sara van de Geer
Prof. Dr. Martin Larsson

Serie 8

April 28th, 2015

Q1. Let X_1 and X_2 follow a normal distribution with mean μ_i and variance σ_i^2. Prove that if X_1 is independent of X_2 then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Solution: Let's suppose, first, that $\mu_1 = \mu_2 = 0$. We just have to use the convolution formula:

$$
\begin{aligned}
f_{X_1+X_2}(x) &= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma_1^2}} e^{-\frac{(x-y)^2}{2\sigma_2^2}}\,dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} \exp\left(-\frac{\sigma_2^2 y^2 + \sigma_1^2(x-y)^2}{2\sigma_1^2\sigma_2^2}\right) dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} \exp\left(-\frac{\sigma_2^2 y^2 + \sigma_1^2 x^2 + \sigma_1^2 y^2 - 2\sigma_1^2 xy}{2\sigma_1^2\sigma_2^2}\right) dy \\
&= \frac{e^{-\frac{x^2}{2\sigma_2^2}}}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} \exp\left(-(\sigma_1^2+\sigma_2^2)\frac{y^2 - \frac{2\sigma_1^2 xy}{(\sigma_1^2+\sigma_2^2)}}{2\sigma_1^2\sigma_2^2}\right) dy \\
&= \frac{e^{-\frac{x^2}{2\sigma_2^2}+\frac{\sigma_1^2 x^2}{2\sigma_2^2(\sigma_1^2+\sigma_2^2)}}}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} \exp\left(-\frac{\left(y - \frac{\sigma_1^2 x}{(\sigma_1^2+\sigma_2^2)}\right)^2}{\frac{2\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}\right) dy \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}} e^{-\frac{x^2}{2(\sigma_1^2+\sigma_2^2)}},
\end{aligned}
$$

that is the distribution function of a normal random variable with parameter $N(0, \sqrt{\sigma_1^2 + \sigma_2^2})$. For the general case note, that $X_i - \mu_i$ is distributed as $N(0, \sigma_i^2)$. So $(X_1 - \mu_1) + (X_2 - \mu_2) \sim N(0, \sigma_1^2 + \sigma_2^2)$, then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Q2. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(Z_n)_{n\in\mathbb{N}}$ a sequence of random variables.

(a) Prove that if $Z_n \xrightarrow{\mathbb{P}} c \in \mathbb{R}$, then for all bounded and continuous functions f

$$\mathbb{E}\left(f(Z_n)\right) \to f(c).$$

(b) Show that if $Z_n \to c \in \mathbb{R}$ in distribution, then $Z_n \xrightarrow{\mathbb{P}} c$.

Solution:

(a) Take $\epsilon > 0$, we know that there exists $\delta > 0$ so that for all $x \in [c-\delta, c+\delta]$, $|f(x) - f(c)| \leq \epsilon$. Then

$$
\begin{aligned}
|\mathbb{E}\left(f(Z_n) - f(c)\right)| &\leq \mathbb{E}\left(|f(Z_n) - f(c)|\right) \\
&\leq \mathbb{E}\left(|f(Z_n) - f(c)|\mathbf{1}_{\{|Z_n - c| \leq \delta\}}\right) + \mathbb{E}\left(|f(Z_n) - f(c)|\mathbf{1}_{\{|Z_n - c| > \delta\}}\right) \\
&\leq \epsilon + \|f\|_\infty \mathbb{P}(|Z_n - c| > \delta) \to \epsilon.
\end{aligned}
$$

.

(b) Take $\epsilon > 0$ and define

$$f_\epsilon(x) \mapsto \min\left\{\frac{1}{\epsilon}d(x,[c-\epsilon,c+\epsilon]),1\right\}.$$

f_ϵ is clearly a continuous function. Note that $f_\epsilon(x)=0$ if $x\in[c-\epsilon,c+\epsilon]$ and $f(x)=1$ if $|x-c|\geq 2\epsilon$. Then we have that:

$$\mathbb{P}(|X_n-c|\geq 2\epsilon)\leq f_\epsilon(X_n)\to f_\epsilon(c)=0.$$

Q3. Take X_n i.i.d random variable so that

$$\mathbb{E}\left(X_1\right)=1, \quad Var\left(X_1\right)=2,$$

and define $S_n := \sum_{i=1}^n X_i$.

(a) Use Chebyshev-inequality to estimate

$$\mathbb{P}\left(\left|\frac{S_n}{n}-1\right|\leq 0.5\right).$$

What is the value of the bound when $n=40$.

(b) Use the Central Limit Theorem to estimate

$$\mathbb{P}\left(\left|\frac{S_n}{n}-1\right|\leq 0.5\right).$$

What is the value of the bound when $n=40$.

Solution:

(a) We have by Chebyshev inequality that

$$\begin{aligned}\mathbb{P}\left(\left|\frac{S_n}{n}-1\right|\leq 0.5\right) &= 1-\mathbb{P}\left(\left|\frac{S_n}{n}-1\right|\geq 0.5\right)\\ &\geq 1-\frac{\mathbb{E}\left(\left(\frac{S_n}{n}-1\right)^2\right)}{0.25}\\ &= 1-\frac{\mathbb{E}\left(Var(\frac{S_n}{n})\right)}{0.25}\\ &= 1-\frac{8}{n}.\end{aligned}$$

When $n=40$, the bound is 0.8.

(b) Thanks to Central Limit Theorem, we have that:

$$\frac{\sqrt{n}}{\sqrt{Var(X_i)}}\left(\frac{S_n}{n}-\mathbb{E}(X_i)\right)\overset{(d)}{\to} N(0,1).$$

Then, using this property we have

$$\begin{aligned}\mathbb{P}\left(\left|\frac{S_n}{n}-1\right|\leq 0.5\right) &= \mathbb{P}\left(\frac{\sqrt{n}}{\sqrt{2}}\left|\frac{S_n}{n}-1\right|\leq\frac{\sqrt{n}}{\sqrt{2}}0.5\right)\\ &\approx \mathbb{P}(-\sqrt{5}\leq N(0,1)\leq\sqrt{5})\\ &= \phi\left(\sqrt{5}\right)-\phi\left(-\sqrt{5}\right)\approx 0.97.\end{aligned}$$

Q4. Take $x \in [0,1]$. We say that x is normal if for x in its binary form,

$$x = \sum_{n \in \mathbb{N}} x_n 2^{-n} \qquad x_n \in \{0,1\},$$

we have that $\lim_{n\to\infty} \frac{|\{1 \le k \le n : x_k = 1\}|}{n} = \frac{1}{2}$.

(a) Prove that if we have a sequence $(U_n)_{n\in\mathbb{N}}$ i.i.d. Bernoulli with parameter $\frac{1}{2}$, then $U = \sum_{n\in\mathbb{N}} U_n 2^{-n}$ is an uniform random variable in $[0,1]$

(b) Prove that if $U \sim U(0,1)$, $\mathbb{P}(U \text{ is normal}) = 1$.

Solution

(a) First we have to prove that U is measurable. For this we just have to realize that $U^{(m)} := \sum_{n=1}^{m} U_n 2^{-n}$ is measurable because it's the finite sum of measurable function and we have that

$$U_n \to U.$$

Second we have to understand the measure that U produces in $\mathbb{R}$. For this it's enough to show that the measure induced by U coincide with the uniform measure in the intervals of the form

$$\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right],$$

for $k \in \mathbb{N} \in [0, 2^n - 1]$. This is because they generate the Borel σ-algebra.
Note that

$$\mathbb{P}((\exists n \in \mathbb{N})(\forall m \ge n) X_m = 1) = 0,$$

thanks to Borel-Cantelli Lemma. So we can work in

$$\tilde{\Omega} := \Omega \backslash \{\omega \in \Omega : (\exists n \in \mathbb{N})(\forall m \ge n) X_m = 1\},$$

i.e. our probability space is $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}})$ where $\tilde{\mathcal{A}} = \mathcal{A}\mid_{\tilde{\Omega}}$ and $\tilde{\mathbb{P}} := \mathbb{P}\mid_{\tilde{\Omega}}$. Now we have that if $k = \sum_{i=0}^{n-1} k_i 2^n \in \{0,1,..,2^n-1\}$:

$$\begin{aligned}
\mathbb{P}\left(U \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right) &= \mathbb{P}\left(\bigcap_{i=0}^{2^n-1} \{U_{i+1} = k_{n-i}\}\right) \\
&= \prod_{i=0}^{2^n-1} \mathbb{P}(U_{i+1} = k_{n-i}) \\
&= 2^{-n}.
\end{aligned}$$

That is the probability of a uniform random variable to be in $\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right]$.

(b) Take $(U_n)_{n\in\mathbb{N}}$ Bernoulli $\frac{1}{2}$ i.i.d. Thanks to part (a) we have that $U := \sum_{n=1}^{n} U_n 2^{-n}$ is uniform distributed and it's normal iff $\frac{\sum_{k=1}^{n} \mathbf{1}_{\{U_k=1\}}}{n} \to \frac{1}{2}$. Then:

$$\mathbb{P}(U \text{is normal}) = \mathbb{P}\left(\frac{\sum_{k=1}^{n} \mathbf{1}_{\{U_k=1\}}}{n} = 1\right) = 1.$$

Where in the last equality we have used the law of large numbers.