Probabilities and statistics

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

Serie 7

April 21st, 2015

Q1. Let X be a normal random variable.

(a) Prove that if we take $Y := X^2$, then $f_Y(y) = ce^{-\frac{y}{2}}\sqrt{y}\mathbf{1}_{\{y\geq 0\}}$ (We say that Y is distributed according to a χ-squared with one degree of freedom).

(b) If Y_1 and Y_2 are two independent copies of Y, prove that $f_{Y_1+Y_2} = c_2 e^{-\frac{x}{2}}\mathbf{1}_{\{x\geq 0\}}$. What is the name of this distribution.

(c) With the help of induction prove that $\sum_{i=1}^n Y_i$, where $(Y_i)_{i=1}^n$ are independent copies of Y, has as a density function

$$f_{\sum_{i=1}^n Y_i}(x) = c_n x\frac{n}{2} - 1e^{-\frac{x}{2}}\mathbf{1}_{\{x\geq 0\}}.$$

This is call a χ-squared distribution with n degrees of freedom.

Solution:

(a) We have that the CDF of Y for $y \geq 0$ is given by:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = 2F_X(\sqrt{y}) - 1.$$

Then, taking the derivative we have:

$$f_Y(y) = f_x(\sqrt{y})y^{-\frac{1}{2}}\mathbf{1}_{\{y\geq 0\}} = ce^{\frac{y}{2}}y^{-\frac{1}{2}}\mathbf{1}_{\{y\geq 0\}}.$$

(b) By the convolution formula we have that:

$$\begin{aligned}
f_{Y_1+Y_2}(x) &= \int_0^x f_Y(x-y)f_Y(y)dy\mathbf{1}_{\{x\geq 0\}} \\
&= c_1^2\int_0^x (x-y)^{-\frac{1}{2}}e^{-\frac{x-y}{2}}y^{-\frac{1}{2}}e^{-\frac{y}{2}}dy\mathbf{1}_{\{x\geq 0\}} \\
&= c_1^2 e^{-\frac{x}{2}}\int_0^x (x-y)^{-\frac{1}{2}}y^{-\frac{1}{2}}dy\mathbf{1}_{\{x\geq 0\}} \\
&= c_1^2\left(\int_0^1 x(x-ux)^{-\frac{1}{2}}(ux)^{-\frac{1}{2}}du\right)e^{-\frac{x}{2}}\mathbf{1}_{\{x\geq 0\}} \\
&= \left(c_1^2\int_0^1 (1-u)^{-\frac{1}{2}}u^{-\frac{1}{2}}du\right)e^{-\frac{x}{2}}\mathbf{1}_{\{x\geq 0\}}.
\end{aligned}$$

This distribution is that of an exponential random variable.

(c) It's clear that the base case is true, now let's prove the inductive step. Suppose that the proposition is true for $n-1$, then

$$\begin{aligned} f_{\sum_{i=1}^n Y_i}(x) &= \int_0^x f_Y(x-y) f_{\sum_{i=1}^{n-1} Y_i}(y) dy \mathbf{1}_{\{x\geq 0\}} \\ &= c_1 c_{n-1} \int_0^x (x-y)^{-\frac{1}{2}} e^{-\frac{x-y}{2}} y^{\frac{n-1}{2}-1} e^{-\frac{y}{2}} dy \mathbf{1}_{\{x\geq 0\}} \\ &= c_1 c_{n-1} e^{-\frac{x}{2}} \int_0^1 (x-xu)^{-\frac{1}{2}} (xu)^{\frac{n-1}{2}-1} x du \mathbf{1}_{\{x\geq 0\}} \\ &= \left(c_1 c_{n-1} \int_0^1 (1-u)^{-\frac{1}{2}} (u)^{-\frac{n-1}{2}-1} du \right) e^{-\frac{x}{2}} x^{\frac{n}{2}-1} \mathbf{1}_{\{x\geq 0\}}. \end{aligned}$$

Q2. Take the following probability space $(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda|_{[0,1]})$, where $\lambda|_{[0,1]}$ is the Lebesgue measure over $[0,1]$. Let $X_n(\omega) = \mathbf{1}_{A_n}(\omega)$ a sequence of random variables with $A_n \in \mathcal{B}([0,1])$.

(a) Under which condition for $(A_n)_{n\in\mathbb{N}}$ we have that $X_n \xrightarrow{\mathbb{P}} 0$.

(b) Write the event $\{\omega : X_n(\omega) \to 0\}$ with help of the sets $(A_n)_{n\in\mathbb{N}}$.

(c) Find a sequence $(A_n)_{n\in\mathbb{N}}$ of events so that $X_n \xrightarrow{\mathbb{P}} 0$ but $\{\omega : X_n(\omega) \to 0\} = \emptyset$.

Solution:

(a) We know that for all $\epsilon \leq \frac{1}{2}$

$$\mathbb{P}(|X_n| \leq \epsilon) = \mathbb{P}(|X_n| = 0) = \mathbb{P}(A_n^c),$$

so $X_n \xrightarrow{\mathbb{P}} 0$ iff $\mathbb{P}(A_n^c) \to 1$.

(b) Given that X_n takes only values in $\{0,1\}$ we know it converges if from a point onward it only takes the value 0, so

$$\{\omega : \lim X_n(\omega) = 0\} = \bigcup_{k\in\mathbb{N}} \bigcap_{n\geq k} A_n^c = \liminf A_n^c.$$

(c) For $n \in \mathbb{N}$ define $r_n = \lfloor \log_2(n) \rfloor$ and define $k_n = n - 2^{r_n}$. Take

$$A_n = \left[\frac{k_n}{2^{r_n}}, \frac{k_n+1}{2^{r_n}} \right],$$

note that $\mathbb{P}(A_n) = r_n \to 0$, so $X_n \xrightarrow{\mathbb{P}} 0$. Additionally note that for each r_n there are $2^{r_n+1} - 2^{r_n} = 2^{r_n}$ different k_n associated to it and also that:

$$\mathbb{P}\left(\bigcup_{n:r_n=r} A_n \right) = 2^{r_n} \frac{1}{2^{r_n}} = 1,$$

so $\bigcup_{n:r_n=r} A_n = [0,1]$. Then we know that for each $r \in \mathbb{N}$ and for all $x \in [0,1]$ there exits $n \in \mathbb{N}$ so that $r_n = r$ and $x \in A_n$, so $X_n(x)$ is 1 infinitely many times. Thus, $\{\omega : X_n(\omega) \to 0\} = \emptyset$.

Q3. Let $(X_i)_{i\geq 1}$ be a sequence of random variables with

$$\begin{aligned}&\mathbb{E}\,(X_i) = \mu \quad \forall i,\\ &Var(X_i) = \sigma^2 < \infty \quad \forall i,\\ &Cov(X_i, X_j) = R(|i-j|) \quad \forall i, j.\end{aligned}$$

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

(a) Prove that if $\lim_{k\to\infty} R(k) = 0$ then $\lim_{n\to\infty} \frac{S_n}{n} = \mu$ in probability.

(b) Prove that if $\sum_{k\in\mathbb{N}} |R(k)| < \infty$ then $\lim_{n\to\infty} nVar(\frac{S_n}{n})$ exists.

Solution:

(a) Thanks to Čebyšev inequality

$$P\left[\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2} Var\left(\frac{S_n}{n}\right)$$

it's enough to prove that $Var(\frac{S_n}{n}) \to 0\,(n\to\infty)$.
Computing the variance we have:

$$\begin{aligned}Var\left(\frac{S_n}{n}\right) &= Var\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\\ &= \frac{1}{n^2}\left(\sum_{i=1}^n Var(X_i) + 2\sum_{i<j}\mathrm{Cov}(X_i, X_j)\right)\\ &= \frac{1}{n^2}\left(n\sigma^2 + 2\sum_{k=1}^{n-1}(n-k)\,R(k)\right)\\ &= \frac{1}{n}\left(\sigma^2 + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)R(k)\right)\end{aligned}$$

Then it's enough to prove that:

$$\lim_{n\to\infty}\frac{2}{n}\sum_{k=1}^{n-1}\Big(\frac{n-k}{n}\Big)R(k) = 0\,.$$

Thanks to Schwarz inequality:

$$\forall i, j \quad |R(|i-j|)| = |\mathrm{Cov}[X_i, X_j]| \leq \sqrt{Var[X_i]}\sqrt{Var[X_j]} = \sigma^2 < \infty.$$

Take $\eta > 0$. Then there exist $k_0 \in \mathbb{N}$, so that $|R(k)| < \eta$ for all $k > k_0$. Thus:

$$\text{i)} \qquad \Big|\frac{2}{n}\sum_{k=1}^{k_0}\Big(\frac{n-k}{n}\Big)R(k)\Big| \leq \frac{2k_0}{n}\sigma^2 < \eta\,, \text{ for } n \text{ sufficiently big}$$

$$\text{ii)} \quad \Big|\frac{2}{n}\sum_{k=k_0+1}^{n-1}\Big(\frac{n-k}{n}\Big)R(k)\Big| \leq \frac{2}{n}\sum_{k=k_0+1}^{n-1}\frac{n-k}{n}\,\eta \leq \frac{2}{n}\,\eta\sum_{k=0}^{n-1}\underbrace{\frac{n-k}{n}}_{\leq 1} \leq 2\eta.$$

$$\text{Then,} \quad \Big|\frac{2}{n}\sum_{k=1}^{n-1}\Big(\frac{n-k}{n}\Big)R(k)\Big| < 3\eta \qquad \forall\; n > n_o = \frac{2k_0\sigma^2}{\eta}\,.$$

In conclusion $\lim_{n\to\infty}\frac{2}{n}\sum_{k=1}^{n-1}\Big(\frac{n-k}{n}\Big)R(k) = 0$. So $\lim_{n\to\infty}\frac{S_n}{n} = \mu$ in probability.

(b) We just have to compute

$$\begin{aligned}\lim_{n\to\infty} n\,Var\left(\frac{S_n}{n}\right) &= \lim_{n\to\infty}\left(\sigma^2 + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)R(k)\right)\\ &= \sigma^2 + 2\sum_{k=1}^{\infty}R(k) - 2\lim_{n\to\infty}\sum_{k=1}^{n-1}\frac{k}{n}\,R(k)\,.\end{aligned}$$

Define:

$$a_n(k) := \begin{cases} \frac{k}{n}\,R(k) & (k < n)\\ 0 & (k \geq n)\end{cases}$$

it's clear that $a_n(k) \to 0$ $(n \to \infty)$ for all k. Then we just have to use dominated convergence to prove that this part goes to 0. Note that $|a_n(k)| \leq |R(k)|$and $|R(k)|$ is absolutely convergente. So:

$$\lim_{n\to\infty}\sum_{k=1}^{n-1}\frac{k}{n}\,R(k) = \lim_{n\to\infty}\sum_{k=1}^{n-1}a_n(k) = \lim_{n\to\infty}\sum_{k=1}^{\infty}a_n(k) = \sum_{k=1}^{\infty}\lim_{n\to\infty}a_n(k) = 0$$

Then

$$\lim_{n\to\infty} n\,Var\Big(\frac{S_n}{n}\Big) = \sigma^2 + 2\sum_{k=1}^{\infty}R(k).$$

Q4. Compute the limit of $\lim_{n\to\infty} e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}$
Hint: You can use the central limit theorem (Skript Theorem 4.3) for $(X_i)_{i\in\mathbb{N}}$ i.d.d. random variables such that $X_i \sim Poi(1)$.

Solution

If we define $S_n := \sum_{i=1}^{n}X_i \sim Poi(n)$, we have that:

$$e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} = \mathbb{P}(S_n \leq n) = \mathbb{P}(S_n \leq n\mathbb{E}\,(X_1)) = \mathbb{P}\left(\frac{1}{\sqrt{n}}\,(S_n - n\mathbb{E}\,(X_1) \leq 0)\right) \to \frac{1}{2}.$$