Exercises for the Stochastic part of Mathematical Data Science WI4231

("Big Data")

Please hand in your answers to 2 of the following 3 exercises.

1. When U and U' are two i.i.d. uniformly distributed random variables on $[0,1]$, show that:
   (a) $\mathrm{E}\left((U-U')^2\right) = \frac{1}{6}$.
   (b) $\mathrm{Var}\left((U-U')^2\right) \approx 0.04$
2. In Giraud's book it is claimed that a p-dimensional ball with radius $r > 0$, one has that the volume $V_p(r)$ of this ball satisfies:
   $$V_p(r) = \frac{\pi^{p/2}}{\Gamma(p/2+1)} r^p \approx \left(\frac{2\pi \mathrm{e} r^2}{p}\right)^{p/2} \frac{1}{\sqrt{p\pi}}, \qquad \text{for large } p. \tag{1}$$
   where Γ is the famous "Gamma-function," defined by:
   $$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\mathrm{e}^{-x}\,\mathrm{d}t, \qquad \text{for } \alpha > 0.$$
   We'er going to show this in a number of steps.
   (a) Show that for the Gamma-function we have that $\Gamma(1) = 1$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, and $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ for all $\alpha > 0$. Use this to show that
   $$\Gamma(p+1) = p! \quad \text{en} \quad \Gamma(p+3/2) = \frac{(2p+1)(2p-1)\cdots 1}{2^{p+1}}\sqrt{\pi},$$
   for $p \in \mathbb{N}$.
   (b) Prove that $V_p(r) = r^p V_p(1)$ for any $p \geq 1$ check that $V_1(1) = 2$ en $V_2(1) = \pi$.
   (c) For $p \geq 3$, show that
   $$\begin{aligned} V_p(1) &= \int_{x_1^2+x_2^2\leq 1} V_{p-2}\left(\sqrt{1-x_2^2-x_2^2}\right)\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= V_{p-2}(1)\int_{r=0}^{r=1}\int_{\theta=0}^{\theta=2\pi}(1-r^2)^{p/2-1} r\,\mathrm{d}r\,\mathrm{d}\theta \\ &= \frac{2\pi}{p}V_{p-2}(1). \end{aligned}$$
   (d) Conclude that
   $$V_{2p}(1) = \frac{\pi^p}{p!} \quad \text{en} \quad V_{2p+1}(1) = \frac{2^{p+1}\pi^p}{(2p+1)(2p-1)\cdots 3}.$$

(e) Use the Stirling expansion

$$\Gamma(\alpha)\alpha^{\alpha-1/2}\mathrm{e}^{-\alpha}\sqrt{2\pi}\left(1+\mathcal{O}(\alpha^{-1})\right) \quad \text{for } \alpha \to +\infty$$

to prove (1).

3. Let Z be a random variable with a standard normal $N(0,1)$-distribution.

(a) For $z > 0$, prove (with integration by parts) that

$$\begin{aligned} \mathrm{P}(|Z| \geq z) &= \sqrt{\frac{2}{\pi}}\frac{\mathrm{e}^{-z^2/2}}{z} - \sqrt{\frac{2}{\pi}}\int_z^\infty x^{-2}\mathrm{e}^{-x^2/2}\,\mathrm{d}x \\ &= \sqrt{\frac{2}{\pi}}\frac{\mathrm{e}^{-z^2/2}}{z}\left(1+\mathcal{O}\left(\frac{1}{z^2}\right)\right). \end{aligned}$$

(b) For $Z_1, \ldots, Z_p$ i.i.d. with $N(0,1)$ standard Gaussian distribution and $\alpha > 0$, show that when $p \to \infty$

$$\begin{aligned} &\mathrm{P}\left(\max_{j=1,\ldots,p}|Z_j| \geq \sqrt{\alpha\log(p)}\right) \\ &= 1-\left(1-\mathrm{P}\left(|Z_1| \geq \sqrt{\alpha\log(p)}\right)\right)^p \\ &= 1-\exp\left(-\sqrt{\frac{2}{\alpha\pi}}\frac{p^{1-\alpha/2}}{\sqrt{\log p}}+\mathcal{O}\left(\frac{p^{1-\alpha/2}}{(\log p)^{3/2}}\right)\right). \end{aligned}$$