Moment and Chernoff bounds

Chapters 2.1,2.2,2.3 from "Concentration Inequalities"

Vardis Kandiros

October 2, 2018

What we study

- We study concentration inequalities, which bound the probability that a real-valued random variable Z differs from its expected value by more than a certain amount.
- More precisely, we seek upper bounds for tail probabilities of the form

$$\mathbf{P}[Z - \mathbf{E}[Z] \geq t] \text{ and } \mathbf{P}[Z - \mathbf{E}[Z] \leq -t]$$

  for $t > 0$.
- We assume that $\mathbf{E}[Z]$ exists.

Markov's Inequality

- Let Y be a non negative random variable. We define $\mathbb{1}_{\{Y \geq t\}}$ to be a random variable that takes value 1 if $Y \geq t$, else 0.
- $Y \geq t\mathbb{1}_{\{Y \geq t\}}$
- By taking expectations in both sides of the inequality,

$$\mathbf{E}[Y] \geq t\mathbf{E}\left[\mathbb{1}_{\{Y \geq t\}}\right] = t\mathbf{P}[Y \geq t] \Rightarrow \mathbf{P}[Y \geq t] \leq \frac{\mathbf{E}[Y]}{t}$$

- By taking $Y = |Z - \mathbf{E}[Z]|$, we get the tail inequality

$$\mathbf{P}[|Z - \mathbf{E}[Z]| \geq t] \leq \frac{\mathbf{E}[|Z - \mathbf{E}[Z]|]}{t}$$

Markov's Inequality

- We can obtain sharper estimates by a small modification.
- Let $\phi : \mathbf{R}^+ \mapsto \mathbf{R}^+$ be an increasing function.
- Obviously

$$\mathbf{P}\left[Y \geq t\right] \leq \mathbf{P}\left[\phi(Y) \geq \phi(t)\right] \leq \frac{\mathbf{E}\left[\phi(Y)\right]}{\phi(t)}$$

Markov's Inequality

- We can obtain sharper estimates by a small modification.
- Let $\phi : \mathbf{R}^+ \mapsto \mathbf{R}^+$ be an increasing function.
- Obviously

$$\mathbf{P}[Y \geq t] \leq \mathbf{P}[\phi(Y) \geq \phi(t)] \leq \frac{\mathbf{E}[\phi(Y)]}{\phi(t)}$$

- If $\phi(t) = t^2$ and $Y = |Z - \mathbf{E}[Z]|$, we obtain *Chebyshev's Inequality*.

$$\mathbf{P}[|Z - \mathbf{E}[Z]| \geq t] \leq \frac{Var(Z)}{t^2}$$

,if $Var(Z)$ exists.

Moment Generating Functions

- Usually we will be interested in tail inequalities about a sum of random variables.
- Thus, we want ϕ to be conveniently handled for sums of random variables.

Moment Generating Functions

- Usually we will be interested in tail inequalities about a sum of random variables.
- Thus, we want ϕ to be conveniently handled for sums of random variables.
- A good choice is $\phi(t) = \mathrm{e}^{\lambda t}$, where $\lambda > 0$.
- Markov's inequality then gives

$$\mathbf{P}\left[Z \geq t\right] \leq \frac{\mathbf{E}\left[\mathrm{e}^{\lambda Z}\right]}{\mathrm{e}^{\lambda t}}$$

, if $\mathbf{E}\left[\mathrm{e}^{\lambda Z}\right]$ exists.

Moment Generating Functions

- Usually we will be interested in tail inequalities about a sum of random variables.
- Thus, we want ϕ to be conveniently handled for sums of random variables.
- A good choice is $\phi(t) = \mathrm{e}^{\lambda t}$, where $\lambda > 0$.
- Markov's inequality then gives

$$\mathbf{P}\left[Z \geq t\right] \leq \frac{\mathbf{E}\left[\mathrm{e}^{\lambda Z}\right]}{\mathrm{e}^{\lambda t}}$$

  , if $\mathbf{E}\left[\mathrm{e}^{\lambda Z}\right]$ exists.
- The function $M(\lambda) = \mathbf{E}\left[\mathrm{e}^{\lambda Z}\right]$ is called the *Moment generating function* of random variable Z.

Properties of Moment Generating Functions

[Bil08] Chapter 21,[Dim15] Chapter 13

- Not all functions have a moment generating function, since the integral $\mathbf{E}\left[\mathrm{e}^{\lambda Z}\right]$ might be $+\infty$.

Properties of Moment Generating Functions

[Bil08] Chapter 21,[Dim15] Chapter 13

- Not all functions have a moment generating function, since the integral $\mathbf{E}\left[e^{\lambda Z}\right]$ might be $+\infty$.
- If $M(s)$ exists for some $s > 0$, then it exists for all $\lambda \in [0, s]$.

Properties of Moment Generating Functions

[Bil08] Chapter 21,[Dim15] Chapter 13

- Not all functions have a moment generating function, since the integral $\mathbf{E}\left[e^{\lambda Z}\right]$ might be $+\infty$.
- If $M(s)$ exists for some $s > 0$, then it exists for all $\lambda \in [0, s]$.
- If $M(\lambda)$ exists for $\lambda \in [-s, s]$, then M is infinitely many times differentiable in $[-s, s]$ and it's n-th derivative is

$$M^{(n)}(\lambda) = \mathbf{E}\left[Z^n e^{\lambda Z}\right]$$

Properties of Moment Generating Functions

[Bil08] Chapter 21,[Dim15] Chapter 13

- Not all functions have a moment generating function, since the integral $\mathbf{E}\left[e^{\lambda Z}\right]$ might be $+\infty$.
- If $M(s)$ exists for some $s > 0$, then it exists for all $\lambda \in [0, s]$.
- If $M(\lambda)$ exists for $\lambda \in [-s, s]$, then M is infinitely many times differentiable in $[-s, s]$ and it's n-th derivative is

$$M^{(n)}(\lambda) = \mathbf{E}\left[Z^n e^{\lambda Z}\right]$$

- Since

$$M''(\lambda) = \mathbf{E}\left[Z^2 e^{\lambda Z}\right] \geq 0$$

the function M is convex in it's domain of definition.

Properties of Moment Generating Functions

[Bil08] Chapter 21,[Dim15] Chapter 13

- Not all functions have a moment generating function, since the integral $\mathbf{E}\left[\mathrm{e}^{\lambda Z}\right]$ might be $+\infty$.
- If $M(s)$ exists for some $s > 0$, then it exists for all $\lambda \in [0, s]$.
- If $M(\lambda)$ exists for $\lambda \in [-s, s]$, then M is infinitely many times differentiable in $[-s, s]$ and it's n-th derivative is

$$M^{(n)}(\lambda) = \mathbf{E}\left[Z^n \mathrm{e}^{\lambda Z}\right]$$

- Since

$$M''(\lambda) = \mathbf{E}\left[Z^2 \mathrm{e}^{\lambda Z}\right] \geq 0$$

  the function M is convex in it's domain of definition.
- If $M(\lambda)$ exists for $\lambda \in [-s, s]$, then $\mathbf{E}\left[Z^n\right] < \infty$ for all $n \in \mathbf{Z}$ and

$$M(\lambda) = \sum_{n=0}^{\infty} \frac{\mathbf{E}\left[Z^n\right] \lambda^n}{n!}$$

The Cramer-Chernoff method

- From now on, we will implicitly assume that for all our random variables the generating function is defined in an interval of positive length.
- We have seen that

$$\mathbf{P}\left[Z \geq t\right] \leq \mathrm{e}^{-\lambda t}\mathbf{E}\left[\mathrm{e}^{\lambda Z}\right] \quad (1)$$

so in order to obtain a good upper bound, we optimize the right hand side with respect to λ.

The Cramer-Chernoff method

- From now on, we will implicitly assume that for all our random variables the generating function is defined in an interval of positive length.
- We have seen that

$$\mathbf{P}[Z \geq t] \leq \mathrm{e}^{-\lambda t}\mathbf{E}\left[\mathrm{e}^{\lambda Z}\right] \tag{1}$$

  so in order to obtain a good upper bound, we optimize the right hand side with respect to λ.
- This is equivalent to minimizing the quantity

$$-\lambda t + \psi_Z(\lambda)$$

  where $\psi_Z(\lambda) = \log \mathbf{E}\left[\mathrm{e}^{\lambda Z}\right]$.
- If we define $\psi_Z^*(t) = \sup_{\lambda > 0}(\lambda t - \psi_Z(\lambda))$, then (1) gives

$$\mathbf{P}[Z \geq t] \leq \mathrm{e}^{-\psi_Z^*(t)}$$

  which is *Chernoff's inequality*.

The Cramer-Chernoff method

- The minimum point is found by setting the derivative to 0.

$$t = \psi'_Z(\lambda) \tag{2}$$

- We can show that ψ_Z is strictly convex, which means that ψ'_Z is strictly increasing and thus invertible.
- This means that (2) has a unique solution λ_t.
- We proceed by calculating λ_t for various distributions of Z.

Examples

Example 1 - Gaussian

- Suppose that $Z \sim N(0, \sigma^2)$.
- $\psi_Z(\lambda) = \frac{\lambda^2 \sigma^2}{2}$
- $$\psi'_Z(\lambda_t) = t \Rightarrow \lambda_t = \frac{t}{\sigma^2}$$
  $$\mathbf{P}[Z \geq t] \leq \mathrm{e}^{-\frac{t^2}{2\sigma^2}}$$
- The constant cannot be improved by more than 1/2.

Examples

Example 1 - Gaussian

- Suppose that $Z \sim N(0, \sigma^2)$.
- $\psi_Z(\lambda) = \frac{\lambda^2 \sigma^2}{2}$
- $$\psi'_Z(\lambda_t) = t \Rightarrow \lambda_t = \frac{t}{\sigma^2}$$
  $$\mathbf{P}[Z \geq t] \leq \mathrm{e}^{-\frac{t^2}{2\sigma^2}}$$
- The constant cannot be improved by more than $1/2$.

Example 2 - Poisson

- Let Y be a Poisson with parameter v and $Z = Y - v$ the centered version.
- $$\mathbf{E}\left[\mathrm{e}^{\lambda Z}\right] = \mathrm{e}^{-\lambda v} \sum_{n=0}^{\infty} \frac{\mathrm{e}^{\lambda n} v^n}{n!} = \mathrm{e}^{-\lambda v - v} \mathrm{e}^{v \mathrm{e}^{\lambda}}$$
- Thus
  $$\mathbf{P}[Z \geq t] \leq \mathrm{e}^{-v h(\frac{t}{v})}$$
  where
  $h(x) = (1 + x)\ln(1 + x) - x$,

SubGaussian RV's

[Rig15] Chapter 1

- A lot of interesting distributions have tails decreasing faster than the Normal distribution.
- To formalize this, we say a r.v. X is *Subgaussian* with *variance factor* v if $\ln \mathbf{E}\left[\mathrm{e}^{\lambda X}\right] \leq \frac{\lambda^2 v}{2}$. The collection of these r.v's is $\mathcal{G}(v)$.
- If $X \in \mathcal{G}(v)$ then $\mathbf{P}\left[X \geq t\right] \leq \mathrm{e}^{-\frac{t^2}{2v}}$.

SubGaussian RV's

[Rig15] Chapter 1

- A lot of interesting distributions have tails decreasing faster than the Normal distribution.
- To formalize this, we say a r.v. X is *Subgaussian* with *variance factor* v if $\ln \mathbf{E}\left[e^{\lambda X}\right] \leq \frac{\lambda^2 v}{2}$. The collection of these r.v's is $\mathcal{G}(v)$.
- If $X \in \mathcal{G}(v)$ then $\mathbf{P}[X \geq t] \leq e^{-\frac{t^2}{2v}}$.

Characterisation Theorem

Let X be an r.v. with $\mathbf{E}[X] = 0$. If for some $v > 0$:

$$\mathbf{P}[X \geq t] \vee \mathbf{P}[X \leq -t] \leq e^{-\frac{t^2}{2v}}$$

then

$$\mathbf{E}\left[X^{2q}\right] \leq q!(4v)^q$$

Conversely, if $\mathbf{E}\left[X^{2q}\right] \leq q! C^q$ then $X \in \mathcal{G}(4C)$.

Exercise 2.1

Exercise 2.1

Let $\boldsymbol{M}Z$ be a median of a random variable Z with $\mathbf{E}\left[Z^2\right] < \infty$. This means $\mathbf{P}\left[Z \geq \boldsymbol{M}Z\right] \geq 1/2$ and $\mathbf{P}\left[Z \leq \boldsymbol{M}Z\right] \geq 1/2$. Then $|\boldsymbol{M}Z - \mathbf{E}\left[Z\right]| \leq \sqrt{VarZ}$

Exercise 2.1

Exercise 2.1

Let $\boldsymbol{M}Z$ be a median of a random variable Z with $\mathbf{E}\left[Z^2\right] < \infty$. This means $\mathbf{P}\left[Z \geq \boldsymbol{M}Z\right] \geq 1/2$ and $\mathbf{P}\left[Z \leq \boldsymbol{M}Z\right] \geq 1/2$. Then $|\boldsymbol{M}Z - \mathbf{E}\left[Z\right]| \leq \sqrt{VarZ}$

Proof

- Without loss of generality we can assume that $\boldsymbol{M}Z \geq 0$. If $\boldsymbol{M}Z \leq 0$, then $-\boldsymbol{M}Z$ is a median of $-Z$ and $Var(Z) = Var(-Z)$. Also, we can assume that $\mathbf{E}\left[Z\right] = 0$.

Exercise 2.1

Exercise 2.1

Let $\boldsymbol{M}Z$ be a median of a random variable Z with $\mathbf{E}\left[Z^2\right] < \infty$. This means $\mathbf{P}\left[Z \geq \boldsymbol{M}Z\right] \geq 1/2$ and $\mathbf{P}\left[Z \leq \boldsymbol{M}Z\right] \geq 1/2$. Then $|\boldsymbol{M}Z - \mathbf{E}\left[Z\right]| \leq \sqrt{VarZ}$

Proof

- Without loss of generality we can assume that $\boldsymbol{M}Z \geq 0$. If $\boldsymbol{M}Z \leq 0$, then $-\boldsymbol{M}Z$ is a median of $-Z$ and $Var(Z) = Var(-Z)$. Also, we can assume that $\mathbf{E}\left[Z\right] = 0$.
- Let $a = \boldsymbol{M}Z$. It suffices to prove $\mathbf{E}\left[Z^2\right] \geq a^2$. We have:

$$\begin{aligned}
\mathbf{E}\left[Z^2\right] &= \mathbf{E}\left[Z^2 \mathbb{1}_{\{Z \geq a\}}\right] + \mathbf{E}\left[Z^2 \mathbb{1}_{\{Z < a\}}\right] \\
&\geq a^2 \mathbf{P}\left[Z \geq a\right] + \mathbf{E}\left[Z^2 \mathbb{1}_{\{Z < a\}}\right] \\
&\geq \frac{a^2}{2} + \mathbf{E}\left[Z^2 \mathbb{1}_{\{Z < a\}}\right]
\end{aligned}$$

Exercise 2.1

Proof.

- It suffices to show $\mathbf{E}\left[Z^2\mathbb{1}_{\{Z<a\}}\right] \geq a^2/2$.

Exercise 2.1

Proof.

- It suffices to show $\mathbf{E}\left[Z^2 \mathbb{1}_{\{Z<a\}}\right] \geq a^2/2$.
- $\mathbf{E}[Z] = 0 \Rightarrow \mathbf{E}\left[Z \mathbb{1}_{\{Z<a\}}\right] = -\mathbf{E}\left[Z \mathbb{1}_{\{Z\geq a\}}\right] \Rightarrow \mathbf{E}\left[Z \mathbb{1}_{\{Z<a\}}\right]^2 = \mathbf{E}\left[Z \mathbb{1}_{\{Z\geq a\}}\right]^2$

Exercise 2.1

Proof.

- It suffices to show $\mathbf{E}\left[Z^2 \mathbb{1}_{\{Z<a\}}\right] \geq a^2/2$.
- $\mathbf{E}[Z] = 0 \Rightarrow \mathbf{E}\left[Z\mathbb{1}_{\{Z<a\}}\right] = -\mathbf{E}\left[Z\mathbb{1}_{\{Z\geq a\}}\right] \Rightarrow \mathbf{E}\left[Z\mathbb{1}_{\{Z<a\}}\right]^2 = \mathbf{E}\left[Z\mathbb{1}_{\{Z\geq a\}}\right]^2$
- But $\mathbf{E}\left[Z\mathbb{1}_{\{Z\geq a\}}\right]^2 \geq \frac{a^2}{4}$. Hence,$\mathbf{E}\left[Z\mathbb{1}_{\{Z<a\}}\right]^2 \geq \frac{a^2}{4}$

Exercise 2.1

Proof.

- It suffices to show $\mathbf{E}\left[Z^2 \mathbb{1}_{\{Z<a\}}\right] \geq a^2/2$.
- $\mathbf{E}[Z] = 0 \Rightarrow \mathbf{E}\left[Z\mathbb{1}_{\{Z<a\}}\right] = -\mathbf{E}\left[Z\mathbb{1}_{\{Z\geq a\}}\right] \Rightarrow \mathbf{E}\left[Z\mathbb{1}_{\{Z<a\}}\right]^2 = \mathbf{E}\left[Z\mathbb{1}_{\{Z\geq a\}}\right]^2$
- But $\mathbf{E}\left[Z\mathbb{1}_{\{Z\geq a\}}\right]^2 \geq \frac{a^2}{4}$. Hence,$\mathbf{E}\left[Z\mathbb{1}_{\{Z<a\}}\right]^2 \geq \frac{a^2}{4}$
- By the Cauchy-Shwarz inequality we get

$$\begin{aligned}
\mathbf{E}\left[Z\mathbb{1}_{\{Z<a\}}\right]^2 &= \mathbf{E}\left[Z\mathbb{1}_{\{Z<a\}}\mathbb{1}_{\{Z<a\}}\right]^2 \\
&\leq \mathbf{E}\left[Z^2\mathbb{1}_{\{Z<a\}}\right]\mathbf{E}\left[\mathbb{1}^2_{\{Z<a\}}\right] \\
&= \mathbf{E}\left[Z^2\mathbb{1}_{\{Z<a\}}\right]\mathbf{P}\left[Z<a\right] \\
&\leq \mathbf{E}\left[Z^2\mathbb{1}_{\{Z<a\}}\right]/2
\end{aligned}$$

which proves the claim.

Exercise 2.3

Exercise 2.3

If $\mathbf{E}\left[Y^2\right] < \infty$, then $\mathbf{P}\left[Y - \mathbf{E}\left[Y\right] \geq t\right] \leq \frac{Var(Y)}{Var(Y)+t^2}$

Exercise 2.3

Exercise 2.3

If $\mathbf{E}\left[Y^2\right] < \infty$, then $\mathbf{P}\left[Y - \mathbf{E}\left[Y\right] \geq t\right] \leq \frac{Var(Y)}{Var(Y)+t^2}$

Proof.

Assume without loss of generality that $\mathbf{E}\left[Y\right] = 0$. We consider functions ϕ_u of the form $\phi_u(t) = (t+u)^2$, for $u > 0$. By Markov's inequality $\mathbf{P}\left[Y \geq t\right] \leq \mathbf{P}\left[\phi_u(Y) \geq \phi_u(t)\right] \leq \frac{\mathbf{E}[\phi_u(Y)]}{\phi_u(t)} = \frac{Var(Y)+u^2}{(t+u)^2}$. By setting $u = \frac{Var(Y)}{t}$ we obtain

$$\mathbf{P}\left[Y \geq t\right] \leq \frac{Var(Y) + \left(\frac{Var(Y)}{t}\right)^2}{(t + \frac{Var(Y)}{t})^2} = \frac{Var(Y)}{Var(Y) + t^2}$$

□

Exercise 2.4

Exercise 2.4

If Y is nonnegative and square integrable, and $a \in (0, 1)$,
$\mathbf{P}[Y \geq a\mathbf{E}[Y]] \geq (1-a)^2 \frac{\mathbf{E}[Y]^2}{\mathbf{E}[Y^2]}$

Exercise 2.4

Exercise 2.4

If Y is nonnegative and square integrable, and $a \in (0,1)$,
$\mathbf{P}[Y \geq a\mathbf{E}[Y]] \geq (1-a)^2 \frac{\mathbf{E}[Y]^2}{\mathbf{E}[Y^2]}$

Proof.

Without loss of generality, $\mathbf{E}[Y] = 1$. We have
$\mathbf{P}[Y \geq a] = \mathbf{P}[1 - Y \leq 1 - a] = 1 - \mathbf{P}[1 - Y > 1 - a]$. By applying the Chebyshev-Cantelli inequality we get

$$\mathbf{P}[1 - Y > 1 - a] \leq \frac{Var(1-Y)}{Var(1-Y) + (1-a)^2} = \frac{Var(Y)}{Var(Y) + (1-a)^2}$$

. Consequently $\mathbf{P}[Y \geq a] \geq \frac{(1-a)^2}{Var(Y)+(1-a)^2} \geq \frac{(1-a)^2}{\mathbf{E}[Y^2]}$ □

Exercise 2.5

Exercise 2.5

If Y is nonnegative and $t > 0$, $\inf_{q \in \mathbf{N}} \mathbf{E}\left[Y^q\right] t^{-q} \leq \inf_{\lambda > 0} \mathbf{E}\left[\mathrm{e}^{\lambda(Y-t)}\right]$

Exercise 2.5

Exercise 2.5

If Y is nonnegative and $t > 0$, $\inf_{q \in \mathbf{N}} \mathbf{E}\left[Y^q\right] t^{-q} \leq \inf_{\lambda > 0} \mathbf{E}\left[e^{\lambda(Y-t)}\right]$

Proof.

Let $A = \inf_{q \in \mathbf{N}} \mathbf{E}\left[Y^q\right] t^{-q}$. For every $\lambda > 0$, we have

$$\begin{aligned} \mathbf{E}\left[e^{\lambda(Y-t)}\right] &= e^{-\lambda t} \sum_{q=0}^{\infty} \lambda^q \frac{\mathbf{E}\left[Y^q\right]}{q!} \\ &\geq e^{-\lambda t} \sum_{q=0}^{\infty} \lambda^q \frac{A t^q}{q!} \\ &= A \end{aligned}$$

□

Exercise 2.7

Exercise 2.7

If $Z \sim N(0, \sigma^2)$ then $\sup_{t>0} \mathbf{P}[Z \geq t] \mathrm{e}^{\frac{t^2}{2\sigma^2}} = 1/2$.

Exercise 2.7

Exercise 2.7

If $Z \sim N(0, \sigma^2)$ then $\sup_{t>0} \mathbf{P}[Z \geq t] \mathrm{e}^{\frac{t^2}{2\sigma^2}} = 1/2$.

Proof.

Let $f(t) = \mathbf{P}[Z \geq t] \mathrm{e}^{\frac{t^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{\frac{t^2}{2\sigma^2}} \int_t^\infty \mathrm{e}^{-\frac{x^2}{2\sigma^2}} \mathrm{d}x$. We get:
$f'(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \left(\frac{t}{\sigma^2} \mathrm{e}^{\frac{t^2}{2\sigma^2}} \int_t^\infty \mathrm{e}^{-\frac{x^2}{2\sigma^2}} \mathrm{d}x - 1 \right)$. We now notice that
$\int_t^\infty \mathrm{e}^{-\frac{x^2}{2\sigma^2}} \mathrm{d}x \leq \int_t^\infty \frac{x}{t} \mathrm{e}^{-\frac{x^2}{2\sigma^2}} \mathrm{d}x = \frac{\sigma^2}{t} \mathrm{e}^{-\frac{t^2}{2\sigma^2}}$. Thus $f'(t) \leq 0$ and
$f(t) \leq f(0) = 1/2$. □

Exercise 2.8

$-\ln(1-u) - u \leq \frac{u^2}{2(1-u)}$ for $u \in (0,1)$.

$h(u) = (1+u)\ln(1+u) - u \geq \frac{u^2}{2(1+u/3)}$,for $u > 0$.

Exercise 2.8

$-\ln(1-u) - u \leq \frac{u^2}{2(1-u)}$ for $u \in (0,1)$.
$h(u) = (1+u)\ln(1+u) - u \geq \frac{u^2}{2(1+u/3)}$,for $u > 0$.

Proof.

For the first one, by Taylor's expansion we obtain

$$-\ln(1-u) - u = \sum_{n=2}^{\infty} \frac{u^n}{n} = u^2 \sum_{n=0}^{\infty} \frac{u^n}{n+2} \leq u^2/2 \sum_{n=0}^{\infty} u^n = u^2/2(1-u)$$

For the second,

$$h(u) = \sum_{n=2}^{\infty} (-1)^n \frac{u^n}{n(n-1)} = u^2 \sum_{n=0}^{\infty} (-1)^n \frac{u^n}{(n+1)(n+2)}$$

Now notice that $(n+1)(n+2) \leq 2 \cdot 3^n$ and the result follows. □

Exercise 2.9

Exercise 2.9

If $X \geq 0$ with $\mathbf{E}\left[X^2\right] < \infty$ then for every $\lambda > 0$,

$$\mathbf{E}\left[\mathrm{e}^{-\lambda(X-\mathbf{E}[X])}\right] \leq \mathrm{e}^{\lambda^2\mathbf{E}\left[X^2\right]/2}$$

Exercise 2.9

Exercise 2.9

If $X \geq 0$ with $\mathbf{E}\left[X^2\right] < \infty$ then for every $\lambda > 0$,

$$\mathbf{E}\left[\mathrm{e}^{-\lambda(X-\mathbf{E}[X])}\right] \leq \mathrm{e}^{\lambda^2 \mathbf{E}\left[X^2\right]/2}$$

Proof.

We use the well known inequalities $\mathrm{e}^{-x} \leq 1 - x + x^2/2$, if $x > 0$ and $1 + x \leq \mathrm{e}^x$.

$$\begin{aligned}
\mathbf{E}\left[\mathrm{e}^{-\lambda(X-\mathbf{E}[X])}\right] &\leq \mathrm{e}^{\mathbf{E}[X]}\mathbf{E}\left[1 - \lambda X + \lambda^2 X^2/2\right] \\
&= \mathrm{e}^{\mathbf{E}[X]}(1 - \lambda\mathbf{E}\left[X\right] + \lambda^2\mathbf{E}\left[X^2\right]/2) \\
&\leq \mathrm{e}^{\mathbf{E}[X]}\mathrm{e}^{-\lambda\mathbf{E}[X] + \lambda^2\mathbf{E}\left[X^2\right]/2} \\
&= \mathrm{e}^{\lambda^2\mathbf{E}\left[X^2\right]/2}
\end{aligned}$$

References

Patrick Billingsley.
Probability and measure.
John Wiley & Sons, 2008.

Cheliotis Dimitris.
A second course in probability, 2015.
URL : https://repository.kallipos.gr/handle/11419/2825.

Philippe Rigollet.
Lecture notes in high dimensional statistics, 2015.
URL : https://ocw.mit.edu/courses/mathematics/18-s997-high-dimensional-statistics-spring-2015/lecture-notes/MIT18_S997S15_Chapter1.pdf.