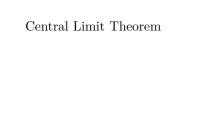
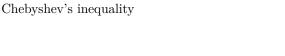
Weak Law of Large Numbers	Let X_i be iid random variables with finite expectation and second moment. Then, for any $\varepsilon > 0$, $\lim_{n \to \infty} \mathbb{P}\left(\left \frac{\sum_{i=1}^n X_i}{n} - \frac{1}{2}\right > \varepsilon\right) = 0$ Proof. By Chebyshev, $\mathbb{P}\left(\left \frac{\sum_{i=1}^n (X_i - \mu)}{n}\right \ge t\right) \le \frac{n\sigma^2}{n^2t^2} = \frac{\sigma^2}{nt^2} \to 0$
weak-law-large-numbers	assuming we have finite variance. \square
Central Limit Theorem	Let X_i be iid random variables with mean μ and variance σ^2 . Then $\frac{\sum_{i=1}^n (X_i - \mu)}{\sigma \sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)$









talagrand-principle

chebyshev-inequality Talagrand's principle

A smooth function of many independent random variables

concentrates around its mean.

 ${\it Proof.} \ \, {\rm By \ Markov},$

For a random variable with mean μ and variance σ^2 ,

 $\mathbb{P}(|X - \mu| \ge t) \le \frac{\sigma^2}{t^2}$

 $\mathbb{P}(|X - \mu| \ge t) = \mathbb{P}((X - \mu)^2 \ge t^2) \le \frac{\sigma^2}{t^2}$

Markov's inequality	Let Y be a nonnegative random variable. Then for all $t>0$ we have $\mathbb{P}(T\geq t)\leq \frac{\mathbb{E}Y}{t}$ Proof. Observe that $Y\geq Y1_{Y\geq t}\geq t1_{Y\geq t}$ and take expectations. \square
log-MGF of a random variable Z	$\psi_Z(\lambda) = \log \mathbb{E} e^{\lambda Z}$
log-mgf log-mgf-def	
Cramer transform	$\psi_Z^*(t) = \sup_{\lambda \ge 0} \lambda t - \phi_Z(\lambda)$
log-mgf cramer-transform cramer-transform-def	
Chernoff bound	$\mathbb{P}(Z \ge t) \le \exp(-\psi_Z^*(t))$
cramer-transform	

Basic properties of ψ_Z and ψ_Z^*	• ψ_Z is infinitely differentiable on $]0, \sup\{\lambda \mid \phi_Z(\lambda) < \infty\}[$ because the MGF is.
	• ψ_Z is convex: If $a, b \ge 0, a + b = 1$, then
	$\mathbb{E}e^{(a\lambda_1+b\lambda_2)Z} = \mathbb{E}(e^{\lambda_1 X})^a (e^{\lambda_2 Z})^b \le (\mathbb{E}e^{\lambda_1 Z})^a (\mathbb{E}e^{\lambda_2 Z})^b$
	by Hölder.
	• ψ_Z^* is nonnegative because $\lambda t - \psi_Z(\lambda) = 0$ when $\lambda = 0$.
	• ψ_Z^* is convex because it is the supremum of linear functions.

log-mgf cramer-transform log-mgf-cramer-transform-properties How to unconstrain ψ_Z^*

log-mgf cramer-transform cramer-transform-unconstrained

MGF and log-MGF of the gaussian distribution

log-mgf log-mgf-gaussian

Cramer transform and Chernoff bound for the gaussian distribution

So $\lambda t - \psi_Z(\lambda) = \lambda t - \frac{\lambda^2 \sigma^2}{2}$ is maximised at $\lambda = \frac{t}{\sigma^2}$ and, for all $t \ge 0$,

 $\psi_Z^*(t) = \sup_{\lambda \geq 0} \lambda t - \frac{\lambda^2 \sigma^2}{2} = \frac{t^2}{2\sigma^2}$ Hence the Chernoff bound is

So the log-MGF is

 $\mathbb{P}(Z \ge t) \le \exp\left(-\frac{t^2}{2\sigma^2}\right)$

If $t > \mathbb{E}Z$ (namely we're looking for a right tail bound), then

 $\psi_Z * = \sup_{\lambda} \lambda t - \psi_Z(\lambda)$

because in general $\mathbb{E}e^{\lambda Z} \geq e^{\lambda \mathbb{E}Z}$ by Jensen, meaning that

 $\lambda t - \psi_Z(\lambda) \le \lambda (t - \mathbb{E}Z) < 0 \le \psi_Z^*(t)$

Complete the square inside the exponent to get

 $\mathbb{E}e^{\lambda Z} = \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} e^{\lambda t} dt$

 $=e^{\frac{\lambda^2\sigma^2}{2}}\int\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(t-\lambda\sigma^2)^2}{2\sigma^2}}\;dt$

 $\psi_Z(\lambda) = \frac{\lambda^2 \sigma^2}{2}$

 $\psi_Z(\lambda) = \frac{\lambda^2 \sigma^2}{2}$

The log-MGF of the gaussian distribution is

 $\psi_Z(\lambda) \geq \lambda \mathbb{E} Z$ and that, if $\lambda < 0$ then

cramer-transform-gaussian

Subgaussian random variables	A random variable X with mean 0 is subgaussian with variance parameter ν if
	$\psi_X(\lambda) \le \frac{\lambda^2 \nu}{2}$
	for all λ . The set of all subgaussian random variables with variance parameter ν is denoted $\mathcal{G}(\nu)$.
subgaussian subgaussian-def	

• If $X \in \mathcal{G}(\nu)$, then $\mathbb{P}(X \ge t), \mathbb{P}(X \le -t) \le e^{-\frac{t^2}{\top}2\nu]}$.

Basic properties of subgaussian random variables

• If $X \in \mathcal{G}(\nu)$, then $\mathbb{P}(X \ge t)$, $\mathbb{P}(X \le -t) \le e^{-\frac{t^2}{\Gamma} 2\nu]}$. • If $X_i \in \mathcal{G}(\nu_i)$ are independent, then $\sum_i X_i \in \mathcal{G}(\sum_i \nu_i)$.

subgaussian subgaussian-basic If $X \in \mathcal{G}(\nu)$, then $\operatorname{Var} X \leq \nu$.

We know

 $\mathbb{E}e^{\lambda X} \le e^{\frac{\lambda^2 \nu}{2}}$

Taylor-expanding and using the fact that $\mathbb{E}X = 0$,

 $1 + \frac{\lambda^2}{2} \mathbb{E} X^2 + O(\lambda^3) \le 1 + \frac{\lambda^2}{2} \nu + O(\lambda^3)$

Taking $\lambda \to 0$,

 $Var X = \mathbb{E}X^2 < \nu$

subgaussian variance subgaussian-variance

Equivalent definitions of subgaussian random variables

The following are equivalent up to choices of ν, b, c, d :

- $X \in \mathcal{G}(\nu)$
- $\forall t > 0, P(X \ge t), \mathbb{P}(X \le -t) \le e^{-\frac{t^2}{2b}}$
- $\forall q, \mathbb{E}X^{2q} \leq q!c^q$
- $\mathbb{E}e^{dX^2} \le 2$

subgaussian-alt