Exam Tips

In this course, N(a, b) means normal distribution with mean a and variance b

pdf of Uniform distribution includes an indicator function!

If $X_n \to X$, $Y_n \to Y$ in distribution, $X_n + Y_n$ may NOT converge in distribution to X + Y

- Slutsky's theorem only guarantee such convergence when Y → c (a constant) in probability.

Method of moment:

- usually used to obtain a rough estimation, used as an initial value of iterative methods.
- estimator of one parameter can depend on another one

Var(y|x) = 0 means that y = f(x) for some function f

sampling distribution: distribution of T(x)

always check that your estimator is genuine: does not depend on θ.

- especially when you define new estimators

Always consider the support of f_θ before proving it is in exponential family

- support should NOT depend on θ

proving a distribution is exponential family: use f = exp(log(f))

Jensen's inequality: if g is convex
$E(g(X)) \geq g(E(X))$

- equality holds when g is linear on a set A with $P(X \in A) = 1$

Sufficiency
Sufficient Statistics may be multi-dimensional (i.e. more than 1 equation)

Bonus for $f_\theta(x|T(x))$ independent of θ when T(x) is sufficient:

- $E_\theta(x|T(x))$ is independent of θ
- CDF is also independent of θ

sufficient statistics is not unique, even minimal sufficient statistics are NOT unique

- counter-example: any bijective function preserves this property

In theorem 1.2: if $f_\theta(x')$ may be 0, use the formulation below: exists function c, c′ s.t.

$$c(x,x')f_\theta(x) \equiv c'(x,x')f_\theta(x') \Leftrightarrow T(x) = T(x')$$

find a sufficient statistics:

- aim to factorise f(x)

prove a given statistics T(x) is sufficient:

- factorise f(x) so that T(x) is used, and f(x) = g(T(x), θ) h(x)

prove a statistics T(x) is NOT sufficient:

- find P(X=x | T(x) = t) and prove it is not free of θ.

Rao-Blackwell estimator using minimal sufficient statistics may NOT give unbiased estimator with the smallest variance.

Likelihood estimation

likelihood equations (derivative = 0) may have multiple solutions, so you still need to check whether the solutions found are global maxima. (second derivative)

- be careful with edge cases

remember the indicator function used when writing likelihood of uniform distribution

- this is one special case where support of f_θ depends on θ
  - You can NOT take log-likelihood in this case, so find MLE using likelihood directly
  - Cramer-Rao lower bound does NOT apply in this case

you can find MLE of g(θ) via chain rule

Asymptotic Behaviour

ways to find asymptotic distribution:

- use central limit theorem
- find CDF/density

- use delta method (suitable for transformations)
- for MLE: use the theorem (but requires: open parameter space, twice continuously differentiable Fisher's information, finite second derivative of Fisher's information, support of pdf does not depend on θ)

Theories of MLE can be used on only one parameter but other parameters should be known

Decision Theory

For square loss function:
- The risk is Variance, if the decision rule is UNBIASED estimator of X
- Bayes estimator is posterior mean

Admissibility is defined from inadmissible, so for proofs with admissibility, contradiction may help

Find a complete and sufficient statistics:
- if the distribution is in exponential family, simply use (T1, ..., Tk)
- First use factorisation theorem, then use definition of completeness

Bayes rule is minimax when its Bayes risk reaches maximum risk. However, Bayes rule has a prior assigned to it whereas minimax does not. So you can modify the parameter of prior if it does not change the Bayes rule.

Hypothesis Testing

For simple assumption $H_0 : \theta = \theta_0$, the level of the set is the size of the set

From the proof of Karlin-Rubin, the theorem also applies to :
$H_0 : \theta = \theta_0$, $H_1 : \theta > \theta_0$
- in this case, you can keep the likelihood ratio test derived from testing $H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1$, (if monotonic likelihood ratio property is satisfied)

When there are two unknown variables μ,σ, if MLE of σ depends on μ, must find MLE of μ first.

(general) Wilk's theorem: distribution of log-likelihood ratio $2\log(t) \sim \chi^2_{m-n}$
- where m is the dimension of parameter space under H_1, m is the dimension of parameter space under H_1