Non-Parametric Statistics Examples

Andrew Tulloch

Contents

CHAPTER 1

Example Sheet 1

Ex. 1 By basic analysis, recall that if $X \sim \textsc{Gamma}(\alpha, 1)$ and $Y \sim \textsc{Gamma}(\beta, 1)$, then

$$\frac{X}{X+Y} \sim \textsc{Beta}(\alpha, \beta) \tag{1.1}$$

Since $S_j \sim \textsc{Gamma}(j, 1)$ and $S_{n+1} - S_j \sim \textsc{Gamma}(n+1-j)$, we have our result.

Now, consider the distribution of $U_{(k)}$. Consider the density $f_{(k)}(x)$ Then we have

$$f_{(k)}(x) = Cx * x^{k-1}(1-x)^{n-k} \tag{1.2}$$

which is of the form of a $\textsc{Beta}(k, n-k+1)$ distribution as required.

Ex. 2 (i) $f(x) = e^{tx}$ is convex on $[a, b]$. So, by definition,

$$f(ta + (1-t)b) \leq tf(a) + (1-t)f(b) \tag{1.3}$$

for all $t \in [0, 1]$.

Then letting $x = ta + (1-t)b$, we have

$$f(x) \leq \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b) \tag{1.4}$$

and taking expectations yields

$$\mathbb{E}(f(x)) \leq \frac{b}{b-a} e^{ta} + \frac{-a}{b-a} e^{tb} \tag{1.5}$$

Letting $p = -\frac{a}{b-a}$, we have $1 - p = \frac{b}{b-a}$, and so ...

Finish writing thi
section up

(ii) We have

$$\begin{aligned}
\mathbb{P}\Big(\sum Y_i > \epsilon\Big) &= \mathbb{P}\Big(e^{t\sum Y_i} > e^{t\epsilon}\Big) && (1.6)\\
&\leq \frac{\mathbb{E}(\prod e^{tY_i})}{e^{t\epsilon}} && (1.7)\\
&= \frac{\prod \mathbb{E}(e^{tY_i})}{e^{t\epsilon}} && (1.8)\\
&\leq \frac{\prod e^{t^2(b_i-a_i)^2/8}}{e^{t\epsilon}} && (1.9)\\
&= e^{\frac{t^2}{8}\sum(b_i-a_i)^2 - t\epsilon} && (1.10)
\end{aligned}$$

and letting

$$s = \frac{4t}{\sum (b_i - a_i)^{2e}} \tag{1.11}$$

and using the union bound we obtain our result.

Ex. 3 This follows easily from the previous result.

$$\mathbb{P}\Big(|\hat{P}_n(A) - \mathbb{P}(A)\,| > \epsilon\Big) = \mathbb{P}\Big(|\sum (\mathbb{I}(X_i \in A) - \mathbb{P}(A))| > n\epsilon\Big) \tag{1.12}$$

$$\leq 2e^{-\frac{2(n\epsilon)^2}{\sum 1^2}} \tag{1.13}$$

$$= 2e^{-2n\epsilon^2} \tag{1.14}$$

Ex. 4 (i) Distribution is multinomial (multivariate binomial) with

$$MN(n, F(t_1), F(t_2) - F(t_1), \dots, 1 - F(t_k)) \tag{1.15}$$

(ii) Consider the distribution of $(\hat{F}_n(t_1), \dots, \hat{F}_n(t_k))$. Then from statistics of the multinomial distribution, we have that this has mean $(F(t_1), \dots, F(t_k))$, and covariance $\Sigma_{ii} = nF(t_{i \vee j})(1 - F(t_{i \wedge j}))$.

Thus, by the multivariate CLT, this merely converges $MN(0, \Sigma_{ii})$

Ex. 5 We have $\mathbb{E}(W_t) = 0$ and

$$\mathrm{Cov}(W_t, W_s) = \mathrm{Cov}(B_t - tB_1, B_s - sB_1) \tag{1.16}$$

$$= \mathrm{Cov}(B_t, B_s) + st\mathrm{Cov}(B_1, B_1) - s\mathrm{Cov}(B_t, B_s) - t\mathrm{Cov}(B_t, B_s) \tag{1.17}$$

$$= t \vee s + ts - t(t \vee s) - s(t \vee s) \tag{1.18}$$

$$= (t \vee s)(1 - t \wedge s) \tag{1.19}$$

Ex. 6 (i)

Type up this long computation

(ii)

Type up this long computation

Ex. 7

Type up this long computation - an double-check we take the Taylor expansion about = 0

Ex. 8 This follows quite easily, we have

$$p_b(x) = \mathbb{P}(X_1 \in I_b(x)) \tag{1.20}$$

$$= \int_{t_b(x)}^{t_b(x)+b} f(y)dy \tag{1.21}$$

$$= \int_{t_b(x)}^{t_b(x)+b} f(x) + f'(y-x)(y-x) + O(b^2) \tag{1.22}$$

$$= bf(x) + f'(x)\int_{t_b(x)}^{t_b(x)+b} (y-x)dy + O(b^3) \tag{1.23}$$

$$= bf(x) + \frac{1}{2}f'(x)[b^2 - 2b(x - t_b(x))] + O(b^3) \tag{1.24}$$

Now, we have

$$\mathbb{E}\Big(\tilde{f}_b(x)\Big) = \frac{\sum \mathbb{P}(X_i \in I_b(x))}{nb} = \frac{p_b(x)}{b} \tag{1.25}$$

and

$$\begin{aligned} \mathbb{V}\Big(\tilde{f}_b(x)\Big) &= \frac{1}{n^2b^2} n\mathbb{V}(\mathbb{I}(X_1 \in I_b(x))) && (1.26) \\ &= \frac{1}{nb^2} p_b(x)(1 - p_b(x)) && (1.27) \\ &= \frac{1}{nb}(f(x)) + O(\frac{1}{n}) && (1.28) \end{aligned}$$

Then using the bias variance decomposition,

$$\begin{aligned} &\frac{f(x)}{nb} + \left(\frac{1}{2}f'(x)[b - 2(x - t_b(x))]\right)^2 + O(\frac{1}{n} + b^3) && (1.29) \\ & && (1.30) \end{aligned}$$

which expands out to the correct solution.

Ex. 9 Follows from integrating over all x, Taylor expanding out the integrands after summing over all bins.

Taking derivatives, we obtain

$$-\frac{1}{nb^2} + \frac{2b}{12}R(f') = 0 \tag{1.31}$$

with solution

$$\left(\frac{6}{nR(f')}\right)^{\frac{1}{3}} \tag{1.32}$$

and so our AMISE scales with

$$n^{-\frac{2}{[}3]}R(f)^{\frac{1}{3}} \tag{1.33}$$

Ex. 10 $|f - f_n|$ is dominated by $f + f_n$, which is integrable. Thus, the DCT gives the required result.

Ex. 11

$$h\int_{-\infty}^{\infty}(K_h^2 \star f)(x)dx = h\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{h^2}K^2(\frac{x-y}{h})f(y)dydx \tag{1.34}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2(u)f(x-hu)dudx \tag{1.35}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2(u)f(x-hu)dxdu \tag{1.36}$$
$$= \int_{-\infty}^{\infty}K^2(u)du \tag{1.37}$$
$$= R(K) \tag{1.38}$$

by Fubini, since all terms are non-negative.

Let $\epsilon > 0$. Then

$$\int_{-\infty}^{\infty}K(u)f(x-hu)du - f(x) = \int_{-\infty}^{\infty}K(u)(f(x-hu)-f(x))du \tag{1.39}$$

and taking absolute values gives

$$\int_{-\infty}^{\infty}|K(u)||f(x-hu)du - f(x)| \leq \int_{-\infty}^{\infty}|K(u)|\epsilon \tag{1.40}$$

for some ϵ' and all $n > N_\epsilon$. Then as $|K(u)| = K(u)$, and $\int_{-\infty}^{\infty}K(u)du = 1$, we have

$$|(K_h \star f)(x) - f(x)| < \epsilon \tag{1.41}$$

for all $n > N_\epsilon$. Thus, we obtain the required result.

Assuming otherwise, there would exist some x' and some sequence h_n such that $(K_h \star f)(x') \to \infty$. However, by the previous result, $f(x') \to \infty$, and by the condition of boundedness of the second derivative, f is bounded. By contradiction, no such x' exists.

Thus, we can apply this result to the functions $g_n = (K_h \star f)^2$ which converges pointwise to $f(x)^2$ (by continuous mapping theorem and the previous result). Thus, by the dominated convergence theorem, we have our result,

$$\int_{-\infty}^{\infty}(K_h \star f)^2(x)dx \to \int_{-\infty}^{\infty}f(x)^2dx \tag{1.42}$$

We have

$$\mathbb{V}\Big(\hat{f}_h(x)\Big) = \frac{1}{n}\hat{\mathbb{V}}\frac{1}{h}K(\frac{x-X}{h}) \tag{1.43}$$
$$= \frac{1}{n}\mathbb{E}\bigg(\frac{1}{h^2}K^2(\frac{x-X}{h})\bigg) - \frac{1}{n}\mathbb{E}\bigg(\frac{1}{h}K(\frac{x-X}{h})\bigg)^2 \tag{1.44}$$
$$= \frac{1}{hn}R(K) + O(\frac{1}{n}) \tag{1.45}$$

Ex. 12

This comes from the exact Taylor expansion of $f(x-hu) = f(x) - \int_0^{hu} f''(t)t$, but it's pretty painful

Bibliography