Change-Point Models Under Diminishing Mis-specification

Moulinath Banerjee

Department of Statistics
The University of Michigan
moulib@umich.edu, http://www.stat.lsa.umich.edu/∼moulib

JSM 2010, Vancouver.

Joint work with Rui Song and Michael Kosorok

Introduction
Formulating the problem
Asymptotics

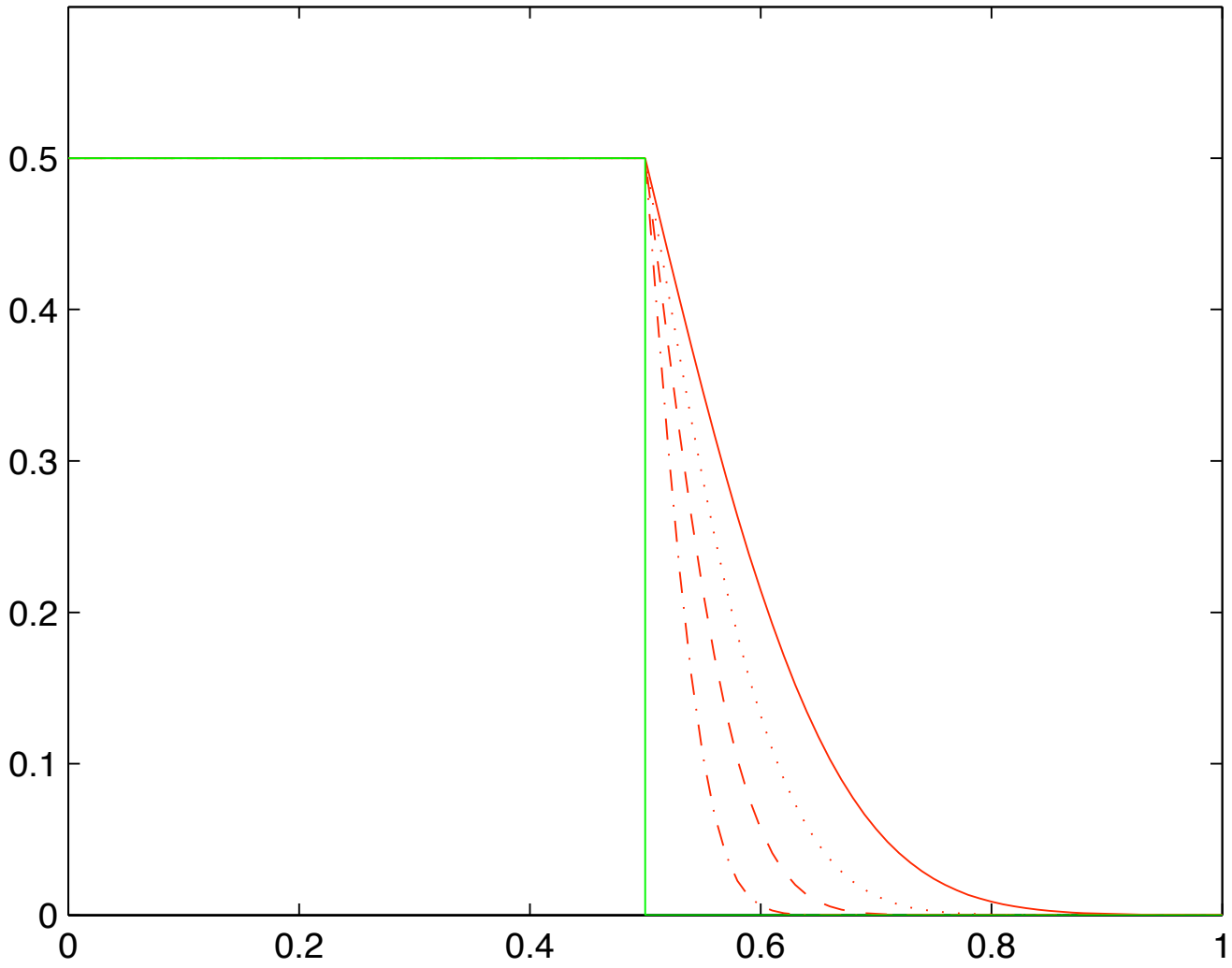
Mis-specified Change Point Model

- Data seem to come from a change-point model when in fact NOT.
- Can happen, for example, if the true regression function rises steeply over a small domain and is relatively flat otherwise and noise is not too large.
- Practitioner would still fit a change-point model.
- How do estimates of the change-point behave under such mis-specification?
- What are intelligent ways of modeling such mis-specification?

Fixed and Local Alternatives

- Think of a postulated change-point model, which for us will be a stump function, as the NULL. Thus, the null model is $\alpha\, 1(x \leq d) + \beta\, 1(x > d)$.
- Mis-specification can then be addressed in terms of either a fixed alternative or local alternatives.
- In the fixed alternative formulation, $Y = f(X) + \epsilon$ where f is a continuous (smooth) function.
- If the true (continuous) f is *close* to a stump function, modeling in terms of local alternatives is attractive (cue from classical parametric theory). In this formulation $Y = f_n(X) + \epsilon$ where f_n is a sequence of functions converging (appropriately) to the NULL.

Sequence of continuous functions converging to a stump



Questions

- What bearing does the speed of convergence of the local alternatives have on the (least-squares) estimates of the change-point?
- What can we say about (a) rates of convergence, and, (b) asymptotic distributions?
- Can we formulate adaptive procedures to estimate the change-point, i.e. procedures that do no necessitate *knowing the degree of mis-specification* in advance?
- Some related work due to Muller et. al., Huskova et.al. on gradually varying change-point models. However, in these settings, the NULL model postulates no change-point. The local alternatives exhibit diminishing jumps.
- Our point of view is the anti-thesis of the above. The situation is clearly not symmetric and different limit distributions obtain.

Fixed alternative

- Fit NULL model $f_{\alpha,\beta,d}(x) \equiv \alpha\, 1(x \leq d) + \beta\, 1(x > d)$ when true model is a fixed alternative $Y = f(X) + \epsilon$, f smooth.

$$(\hat{\alpha}, \hat{\beta}, \hat{d}) = \arg\min \mathbb{P}_n \left[Y - f_{\alpha,\beta,d}(X)\right]^2$$

consistently estimates

$$(\tilde{\alpha}, \tilde{\beta}, \tilde{d}) == \arg\min P \left[Y - f_{\alpha,\beta,d}(X)\right]^2 .$$

- Banerjee and McKeague (2007) show that $n^{1/3}(\hat{d} - \tilde{d})$ converges to a multiple of Chernoff's distribution. Problem also partly addressed by Buhlmann and Yu (2002).
- $\mathbb{Z} \equiv \arg\min_t \; W(t) + t^2$.
- In sharp contrast when the NULL model is true and the true parameters are (α_0, β_0, d^0), $n(\hat{d} - d^0)$ converges to a minimizer of a compound Poisson process. (Kosorok (2008), Ko and Qian (2009))

Formulation

- $Y = f_n(X) + \epsilon$ with $X \sim p_X$ on $[0, 1]$.

$$f_n(x) = f(\alpha_n(x - \theta_0))$$

  with $f : \mathbb{R} \mapsto [0, 1]$, continuously differentiable and $f(-\infty) = 0 = f(\infty) - 1$ and $0 < \theta_0 < 1$.
- Our NULL working model: $f_\theta(x) = \alpha\, 1(x \leq \theta) + \beta\, 1(x > \theta)$.
- α and β are known fixed quantitites with $0 \leq \alpha < \beta \leq 1$.
- We restrict attention to fitting 1 parameter models. The 3 parameter model will be discussed briefly later.

Formulation

- The least squares estimate:

$$\hat{\theta}^n = \arg\min_{\theta} \mathbb{P}_n \left[(Y-\alpha)^2 \, 1(X \leq \theta) + (Y-\beta)^2 \, 1(X > \theta) \right].$$

- Let

$$\mathbb{M}_n(\theta) = \mathbb{P}_n \left[(Y - (\alpha+\beta)/2) \, 1(X \leq \theta) \right].$$

- $\hat{\theta}^n = \arg\min_\theta \mathbb{M}_n(\theta)$ can be viewed as an estimator of $\theta^n = \arg\min_\theta M_n(\theta)$ where

$$M_n(\theta) = P_n \left[(Y - (\alpha+\beta)/2) \, 1(X \leq \theta) \right].$$

- $\theta^n = \theta_0 + \alpha_n^{-1} f^{-1}((\alpha+\beta)/2).$

Rates of convergence

- To determine (a non-trivial) r_n^α such that $r_n^\alpha\,(\hat{\theta}_n - \theta^n)$ is $O_p(1)$.
- Such a rate is determined via the interplay of two different quantities, these being:
  (a) the behavior of $M_n(\theta) - M_n(\theta^n)$ in an *appropriate* neighborhood of θ^n and
  (b) the behavior of the *expected modulus of continuity* of the process $\mathbb{G}_n\, f_n(x, y, \theta)$ in a neighborhood of θ^n.
- Here $\mathbb{G}_n = \sqrt{n}\,(\mathbb{P}_n - P_n)$ and

$$f_n(x, y, \theta) = (y - (\alpha + \beta)/2)\,(1(x \le \theta) - 1(x \le \theta^n))\,.$$

Rates of convergence

- Choose and fix $K > 0$. Under a mild assumption on f,

$$M_n(\theta) - M_n(\theta^n) \geq E_K \, \tilde{d}_n^2(\theta, \theta^n) \,,$$

for all $\tilde{d}_n(\theta, \theta^n) < \delta_0$, where

$$\tilde{d}_n^2(\theta, \theta^n) = \alpha_n \, (\theta - \theta_n)^2 \, 1(|\theta - \theta^n| \leq K/\alpha_n) + |\theta - \theta^n| \, 1(|\theta - \theta^n| > K/\alpha_n) \,.$$

- Via maximal inequalities, for $\delta \leq \delta_0$,

$$E^\star[\sup_{\tilde{d}_n(\theta,\theta^n)<\delta} |\mathbb{G}_n \, (f(x, y, \theta) - f(x, y, \theta^n))| \,] \leq \phi_n(\delta)$$

with

$$\phi_n(\delta) = \frac{\delta^{1/2}}{\alpha_n^{1/4}} \, 1 \left(\delta \leq \frac{1}{\sqrt{\alpha_n}} \right) + \delta \, 1 \left(\delta > \frac{1}{\sqrt{\alpha_n}} \right) \,.$$

Rates of Convergence

- Solving $r_n^2\,\phi_n(1/r_n) \le \sqrt{n}$ yields

$$\frac{r_n^{3/2}}{\alpha_n^{1/4}}\,1(r_n \ge \sqrt{\alpha_n}) + r_n\,1(r_n < \sqrt{\alpha_n}) \le \sqrt{n}\,. \qquad (1)$$

- **Case 1:** $\alpha_n = o(n)$. In this case, we seek a solution with $r_n \ge \sqrt{\alpha_n}$, whence the first term is relevant. A simple calculation shows $r_n = n^{1/3}\,\alpha_n^{1/6}$. Conclude that:

$$n^{1/3}\alpha_n^{1/6}(\alpha_n^{1/2}|\hat{\theta}_n - \theta^n|\,1(\alpha_n|\hat{\theta}_n - \theta^n| \le 1) + |\hat{\theta}_n - \theta^n|^{1/2}\,1(\alpha_n|\hat{\theta}_n - \theta^n| > 1))$$

  is $O_p(1)$. This implies that:

$$n^{1/3}\alpha_n^{2/3}|\hat{\theta}_n - \theta^n|\,1(\alpha_n|\hat{\theta}_n - \theta^n| \le 1) + n^{2/3}\,\alpha_n^{1/3}|\hat{\theta}_n - \theta^n|1(\alpha_n|\hat{\theta}_n - \theta^n| > 1)$$

  is $O_p(1)$.

- Since $n^{1/3}\alpha_n^{2/3}$ is slower than $n^{2/3}\alpha_n^{1/3}$, conclude that $n^{1/3}\alpha_n^{2/3}\,|\hat{\theta}_n - \theta^n| = O_p(1)$.

Rates of convergence

- **Case 2:** For $\alpha_n = n$ both rates $n^{1/3}\,\alpha_n^{2/3}$ and $n^{2/3}\,\alpha_n^{1/3}$ are equal to n and we conclude $n\,|\hat{\theta}_n - \theta^n|$ is $O_p(1)$.

- **Case 3:** For $n = o(\alpha_n)$, the second part in (1) becomes relevant i.e. we seek a solution with $r_n < \sqrt{\alpha}_n$. Conclude that:

$$(n\,\alpha_n)^{1/2}|\hat{\theta}_n - \theta^n|\,1(\alpha_n|\hat{\theta}_n - \theta^n| \leq 1) + n\,|\hat{\theta}_n - \theta^n| 1(\alpha_n|\hat{\theta}_n - \theta^n| > 1)\,.$$

Since $n\,\alpha_n$ is faster than n^2, it follows that $n\,|\hat{\theta}_n - \theta^n|$ is $O_p(1)$. On the other hand, by the observation that the least squares estimate $\hat{\theta}_n$ is atleast as far as θ^n from the X_i closest to the latter and the fact that this X_i converges to θ^n at rate n (in fact, $n\,|X_i - \theta^n|$ converges to an exponential distribution), it follows that n must be the non-trivial rate of convergence.

Asymptotic distribution: $\alpha_n = o(n)$

- Let $\hat{h}_n \equiv n^{1/3}\alpha_n^{2/3}\,(\hat{\theta}^n - \theta^0)$.
- Introducing a local variable and rescaling the criterion function, one gets:

$$\begin{aligned}\hat{h}_n &= \arg\min_h\; n^{2/3}\alpha_n^{1/3}\,\mathbb{P}_n[(Y-(\alpha+\beta)/2)\,(1(X\le\theta^n+h\,n^{-1/3})-1(X\le\theta^n))]\\ &\equiv I_n + II_n\,,\end{aligned}$$

with cb

$$II_n = n^{2/3}\alpha_n^{1/3}\,[M_n(\theta^n + h\,n^{-1/3}\,\alpha_n^{-2/3}) - M_n(\theta^n)]$$

and

$$I_n = \sqrt{n}\,(\mathbb{P}_n - P_n)\,g_{n,h}(x,y)$$

with

$$g_{n,h}(x,y) = n^{1/6}\alpha_n^{1/3}\,[(y-(\alpha+\beta)/2)[1(x\le\theta^n+h\,n^{-1/3})-1(X\le\theta^n))]\,.$$

Next,

$$II_n \to \frac{1}{2}\,h^2\,p_X(\theta_0)\,f'(\lambda_0)\,,$$

and

$$I_n \to_d \sqrt{\sigma^2\,p_X(\theta_0)}\,W(h)\,.$$

- Conclude that $\hat{h}^n \to \arg\min_h\,(\sqrt{\sigma^2\,p_X(\theta_0)}\,W(h) + \frac{1}{2}\,h^2\,p_X(\theta_0)\,f'(\lambda_0))\,.$

Asymptotic distribution: $\alpha_n = n$

- Let $\{\nu^+(h) : h \geq 0\}$ be a homogeneous Poisson process on $[0, \infty)$ with RCLL sample paths and rate parameter $p_X(\theta_0)$.
- Let $\{\epsilon_i\}_{i=1}^{\infty}$ be i.i.d. ϵ and distributed independently of $\nu^+(h)$.
- Let S_i denote the time to the i'th arrival for the Poisson process ν^+, i.e. $S_i = T_1 + T_2 + \ldots + T_i$, where $\{T_j\}_{j=1}^{\infty}$ are the i.i.d. exponential $p_x(\theta_0)$ inter-arrival times corresponding to $\nu^+(h)$.
- For $h \geq 0$, define:

$$\Lambda(h) = \sum_{j=0}^{\nu^+(h)} \left(\epsilon_j + f(\lambda_0 + S_j) - f(\lambda_0)\right) .$$

Asymptotic distribution: $\alpha_n = n$

- To define the process for $h \leq 0$, consider $\nu^-(h)$, an LCRR homogeneous Poisson process on $[0, \infty)$ with parameter $p_X(\theta_0)$ and $\{\tilde{\epsilon}_i\}_{i=1}^{\infty}$ i.i.d. ϵ again, and independent of $\nu^-(h)$.
- Let $\tilde{S}_i$ denote the time to the i'th arrival for the process ν^-.
- For $h \leq 0$, define:

$$\Lambda(h) = \sum_{j=0}^{\nu^-(-h)} \left(-\tilde{\epsilon}_j + f(\lambda_0) - f(\lambda_0 - \tilde{S}_j) \right) .$$

- It can be checked that the process $\Lambda(h)$ thus defined is stationary and has independent increments.
- $V_n(h) \equiv \sum_{i=1}^{n} [Y_i - (\alpha+\beta)/2] [1(X_i \leq \theta^n + h/n) - 1(X_i \leq \theta^n)]$ converges in distribution to $\Lambda(h)$.
- Conclude that $n(\hat{\theta}_n - \theta^n) = \arg\min_h V_n(h) \Rightarrow_d \arg\min_h \Lambda(h)$.

$\alpha_n = n$: DEPENDENCE OF ASYMPTOTICS ON f

- Note that

$$V_n(h) \equiv \sum_{i=1}^{n} [Y_i - (\alpha+\beta)/2]\,[1(X_i \le \theta^n + h/n) - 1(X_i \le \theta^n)]$$

.

- Calculate $E(e^{itV_n(h)}) = \left(1 + \frac{\xi_n}{n}\right)^n$, where

$$\begin{aligned}
\xi_n &= n \int_{\theta^n}^{\theta^n + h/n} [\phi_\epsilon(t)\, \exp[it\{f(n(x-\theta_0)) - f(\lambda_0)\}] - 1]\; p_X(x)\, dx \\
&= \int_0^h [\phi_\epsilon(t)\, \exp[it\{f(n\,(\theta_n + u/n - \theta_0)) - f(\lambda_0)\}] - 1]\; p_X(\theta_0 + u/n)\, du \\
&= \int_0^h [\phi_\epsilon(t)\, \exp[it\{f(\lambda_0 + u) - f(\lambda_0)\}] - 1]\; p_X(\theta_0)\, du + o(1) \\
&\to \phi_\epsilon(t) p_X(\theta_0)\, v(h,t) - p_X(\theta_0)\, h\,,
\end{aligned}$$

where $v(h,t) = \int_0^h \exp[it(f(\lambda_0 + u) - f(\lambda_0))]du\,.$

- Conclude that

$$E(e^{itV_n(h)}) \to \exp\{\phi_\epsilon(t) p_X(\theta_0)\, v(h,t) - p_X(\theta_0)\, h\}\,.$$

$n = o(\alpha_n)$: ASYMPTOTIC DISTRIBUTION

- Define:

$$\tilde{\Lambda}(h) = \sum_{j=0}^{\nu^+(h)} [(1 - (\alpha + \beta)/2) + \epsilon_i]\, 1(h \geq 0)$$

$$+ \sum_{j=0}^{\nu^-(-h)} [(\alpha + \beta)/2 - \tilde{\epsilon}_i]\, 1(h < 0)\,.$$

- Then:

$$\hat{h}_n \equiv n\,(\hat{\theta}_n - \theta^n) = \arg\min_h V_n(h) \Rightarrow_d \arg\min_h \tilde{\Lambda}(h)\,.$$

$n = o(\alpha_n)$: LACK OF DEPENDENCE OF ASYMPTOTICS ON f

- Note that

$$V_n(h) \equiv \sum_{i=1}^{n} [Y_i - (\alpha+\beta)/2]\,[1(X_i \le \theta^n + h/n) - 1(X_i \le \theta^n)]\,.$$

- Calculate $E(e^{itV_n(h)}) = \left(1 + \frac{\xi_n}{n}\right)^n$, where

$$\begin{aligned}
\xi_n &= n \int_{\theta^n}^{\theta^n + h/n} [\phi_\epsilon(t)\, \exp[it\{f(\alpha_n(x-\theta_0)) - f(\lambda_0)\}] - 1]\; p_X(x)\, dx \\
&= \int_0^h [\phi_\epsilon(t)\, \exp[it\{f(\alpha_n(\theta_n + u/n - \theta_0)) - f(\lambda_0)\}] - 1]\; p_X(\theta_0 + u/n)\, du \\
&= \int_0^h [\phi_\epsilon(t)\, \exp[it\{f(\lambda_0 + (\alpha_n u/n)) - f(\lambda_0)\}] - 1]\; p_X(\theta_0)\, du + o(1) \\
&\to \phi_\epsilon(t) p_X(\theta_0)\, v(h,t) - p_X(\theta_0)\, h\,,
\end{aligned}$$

where

$$v(h,t) = \int_0^h \exp[it(1 - f(\lambda_0))] du\; p_X(\theta_0) = \exp[it(1 - (\alpha+\beta)/2))] p_X(\theta_0)\, h\,.$$

- Conclude that

$$E(e^{itV_n(h)}) \to \exp[\phi_\epsilon(t) p_X(\theta_0)\, v(h,t) - p_X(\theta_0)\, h]$$

Bias issues

- For $\alpha_n = o(n)$, the rate of convergence of $\hat{\theta}_n$ for θ^n is $n^{1/3}\alpha_n^{2/3}$ while the rate of convergence of θ^n to θ_0 is slower, namely α_n. So $n^{1/3}\alpha_n^{2/3}(\hat{\theta}_n - \theta_0)$ has an asymptotic bias going to ∞.
- For $\alpha_n = n$ the order of the bias of θ^n for θ_n is the same as the rate of convergence of $\hat{\theta}_n$ to θ^n, namely n, so that $n(\hat{\theta}_n - \theta_0)$ is distributed as $\arg\min_h \Lambda(h) + \lambda_0$.
- For $n = o(\alpha_n)$, the bias reduces at a faster rate than the rate of convergence which is still n and $n(\hat{\theta}_n - \theta_0)$ is distributed as $\arg\min_h \tilde{\Lambda}(h)$.

Asymptotics under the NULL (stump) model

- Suppose that the true model is the NULL model: $f_0(x) = 1(x > \theta_0)$. This is the limit of the models $f_n(x) = f(\alpha_n(x - \theta_0))$.
- Propose $\hat{\theta}_n = \arg\min_\theta \mathbb{P}_n[(Y - (\alpha + \beta)/2)\, 1(X \leq \theta)]$ as an estimate of $\theta_0 = \arg\min_\theta P[(Y - (\alpha + \beta)/2)\, 1(X \leq \theta)]$.
- Not difficult to show that $n(\hat{\theta}_n - \theta_0)$ converges to $\arg\min_h \tilde{\Lambda}(h)$.
- The asymptotic distribution is the exact same as when α_n converges faster than n.

Analogies with parametric models

- Consider regular parametric model $\{p(x, \theta)\}$ and let $X_1, X_2, \ldots, X_n$ i.i.d. θ_{true}.

Analogies with parametric models

- Consider regular parametric model $\{p(x, \theta)\}$ and let $X_1, X_2, \ldots, X_n$ i.i.d. θ_{true}.
- Null hypothesis $\theta = \theta_0$. Remember, our null is the stump model.

Analogies with parametric models

- Consider regular parametric model $\{p(x, \theta)\}$ and let $X_1, X_2, \ldots, X_n$ i.i.d. θ_{true}.
- Null hypothesis $\theta = \theta_0$. Remember, our null is the stump model.
- Under null, $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically $N(0, I(\theta_0)^{-1})$.

Analogies with parametric models

- Consider regular parametric model $\{p(x, \theta)\}$ and let $X_1, X_2, \ldots, X_n$ i.i.d. θ_{true}.
- Null hypothesis $\theta = \theta_0$. Remember, our null is the stump model.
- Under null, $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically $N(0, I(\theta_0)^{-1})$.
- With alternatives converging faster than $\sqrt{n}$, say $\theta_n = \theta_0 + h\, n^{-\gamma}$ with $\gamma > 1/2$, the limit of $\sqrt{n}(\hat{\theta} - \theta_0)$ continues to be the same as under the null.

Analogies with parametric models

- Consider regular parametric model $\{p(x, \theta)\}$ and let $X_1, X_2, \ldots, X_n$ i.i.d. θ_{true}.
- Null hypothesis $\theta = \theta_0$. Remember, our null is the stump model.
- Under null, $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically $N(0, I(\theta_0)^{-1})$.
- With alternatives converging faster than $\sqrt{n}$, say $\theta_n = \theta_0 + h\, n^{-\gamma}$ with $\gamma > 1/2$, the limit of $\sqrt{n}(\hat{\theta} - \theta_0)$ continues to be the same as under the null.
- Similarly, in the change-point model, with alternatives converging faster than n, the rate of convergence under the null, the limit distribution is the same as under the null model.

Analogies with parametric models

- Consider regular parametric model $\{p(x, \theta)\}$ and let $X_1, X_2, \ldots, X_n$ i.i.d. θ_{true}.
- Null hypothesis $\theta = \theta_0$. Remember, our null is the stump model.
- Under null, $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically $N(0, I(\theta_0)^{-1})$.
- With alternatives converging faster than $\sqrt{n}$, say $\theta_n = \theta_0 + h\, n^{-\gamma}$ with $\gamma > 1/2$, the limit of $\sqrt{n}(\hat{\theta} - \theta_0)$ continues to be the same as under the null.
- Similarly, in the change-point model, with alternatives converging faster than n, the rate of convergence under the null, the limit distribution is the same as under the null model.
- In the parametric model, with $\theta_n = \theta_0 + n^{-\gamma}$ and $\gamma < 1/2$, the limit distribution of $\sqrt{n}\,(\hat{\theta}_n - \theta_0)$ has an asymptotic bias going to ∞. As we have seen, a similar thing happens with $\alpha_n = o(n)$ in the change-point model.

Analogies with parametric models

- When $\theta_n = \theta_0 + h/\sqrt{n}$, $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically $N(h, I(\theta_0)^{-1})$.

Analogies with parametric models

- When $\theta_n = \theta_0 + h/\sqrt{n}$, $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically $N(h, I(\theta_0)^{-1})$.
- Thus, with local alternatives converging at the rate of the MLE under the null, the direction of approach of the alternatives figures in the limit distribution.

Analogies with parametric models

- When $\theta_n = \theta_0 + h/\sqrt{n}$, $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically $N(h, I(\theta_0)^{-1})$.
- Thus, with local alternatives converging at the rate of the MLE under the null, the direction of approach of the alternatives figures in the limit distribution.
- Similarly, in the change-point model, when $\alpha_n = n$ (the rate of the change-point estimate under the null model), f, which can be interpreted as the direction of approach of the continuous models to the stump, figures in the limit.

Analogies with parametric models

- When $\theta_n = \theta_0 + h/\sqrt{n}$, $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically $N(h, I(\theta_0)^{-1})$.
- Thus, with local alternatives converging at the rate of the MLE under the null, the direction of approach of the alternatives figures in the limit distribution.
- Similarly, in the change-point model, when $\alpha_n = n$ (the rate of the change-point estimate under the null model), f, which can be interpreted as the direction of approach of the continuous models to the stump, figures in the limit.
- Note that the rate of convergence of the estimator in the change-point model depends on α_n, the rate at which the alternatives approach the null. This is however not the case for parametric models.

Connections between the limit distributions

- Consider $Y = f\left(\alpha_n\left(X - \theta_0\right)\right) + \epsilon$ for an arbitrary α_n.

Connections between the limit distributions

- Consider $Y = f\left(\alpha_n\left(X - \theta_0\right)\right) + \epsilon$ for an arbitrary α_n.
- Write $Y = f_c(n\left(X - \theta_0\right)) + \epsilon$ where $f_c(t) = f(ct)$ and $c = c(n) = \alpha_n/n$.

Connections between the limit distributions

- Consider $Y = f\left(\alpha_n\left(X - \theta_0\right)\right) + \epsilon$ for an arbitrary α_n.
- Write $Y = f_c(n\left(X - \theta_0\right)) + \epsilon$ where $f_c(t) = f(ct)$ and $c = c(n) = \alpha_n/n$.
- Compute $\hat{\theta}_n$ and approximate the distribution of $n(\hat{\theta}_n - \theta_0)$ by that of $\arg\min_h \Lambda_{f_c}(h) + \lambda_{0,f_c}$.

Connections between the limit distributions

- Consider $Y = f\left(\alpha_n\left(X - \theta_0\right)\right) + \epsilon$ for an arbitrary α_n.
- Write $Y = f_c(n\left(X - \theta_0\right)) + \epsilon$ where $f_c(t) = f(ct)$ and $c = c(n) = \alpha_n/n$.
- Compute $\hat{\theta}_n$ and approximate the distribution of $n(\hat{\theta}_n - \theta_0)$ by that of $\arg\min_h \Lambda_{f_c}(h) + \lambda_{0,f_c}$.
- Is this approximation consistent with what we would get if we used the correct limit distribution as an approximation?

Connections between the limit distributions

- Consider $Y = f(\alpha_n(X - \theta_0)) + \epsilon$ for an arbitrary α_n.
- Write $Y = f_c(n(X - \theta_0)) + \epsilon$ where $f_c(t) = f(ct)$ and $c = c(n) = \alpha_n/n$.
- Compute $\hat{\theta}_n$ and approximate the distribution of $n(\hat{\theta}_n - \theta_0)$ by that of $\arg\min_h \Lambda_{f_c}(h) + \lambda_{0,f_c}$.
- Is this approximation consistent with what we would get if we used the correct limit distribution as an approximation?
- Suppose $n = o(\alpha_n)$. Then $c = c(n) \to \infty$ and it appears that (heuristics for now) Λ_{f_c} approaches $\tilde{\Lambda}$. Also, λ_{0,f_c} approaches 0. This is reassuring.

Connections between the limit distributions

- Consider $Y = f(\alpha_n(X - \theta_0)) + \epsilon$ for an arbitrary α_n.
- Write $Y = f_c(n(X - \theta_0)) + \epsilon$ where $f_c(t) = f(ct)$ and $c = c(n) = \alpha_n/n$.
- Compute $\hat{\theta}_n$ and approximate the distribution of $n(\hat{\theta}_n - \theta_0)$ by that of $\arg\min_h \Lambda_{f_c}(h) + \lambda_{0,f_c}$.
- Is this approximation consistent with what we would get if we used the correct limit distribution as an approximation?
- Suppose $n = o(\alpha_n)$. Then $c = c(n) \to \infty$ and it appears that (heuristics for now) Λ_{f_c} approaches $\tilde{\Lambda}$. Also, λ_{0,f_c} approaches 0. This is reassuring.
- When $\alpha_n = o(n)$, calculations reveal that the approximation will work correctly, provided $c_n^{2/3} \arg\min_h \Lambda_{f_{c_n}}(h)$ converges to the limit distribution of $n^{1/3}\alpha_n^{2/3}(\hat{\theta}_n - \theta^n)$. Does this happen? Now $c_n \to 0$.

Connections between the limit distributions

- Consider $Y = f(\alpha_n(X - \theta_0)) + \epsilon$ for an arbitrary α_n.
- Write $Y = f_c(n(X - \theta_0)) + \epsilon$ where $f_c(t) = f(ct)$ and $c = c(n) = \alpha_n/n$.
- Compute $\hat{\theta}_n$ and approximate the distribution of $n(\hat{\theta}_n - \theta_0)$ by that of $\arg\min_h \Lambda_{f_c}(h) + \lambda_{0,f_c}$.
- Is this approximation consistent with what we would get if we used the correct limit distribution as an approximation?
- Suppose $n = o(\alpha_n)$. Then $c = c(n) \to \infty$ and it appears that (heuristics for now) Λ_{f_c} approaches $\tilde{\Lambda}$. Also, λ_{0,f_c} approaches 0. This is reassuring.
- When $\alpha_n = o(n)$, calculations reveal that the approximation will work correctly, provided $c_n^{2/3} \arg\min_h \Lambda_{f_{c_n}}(h)$ converges to the limit distribution of $n^{1/3}\alpha_n^{2/3}(\hat{\theta}_n - \theta^n)$. Does this happen? Now $c_n \to 0$.
- The possibility of conducting inference based on these ideas needs to be investigated.

The three parameter problem

- In this problem, we compute:

$$(\hat{\gamma}_n, \hat{\delta}_n, \hat{\theta}^n) = \arg\min_{\gamma,\delta,\theta} \mathbb{P}_n[(Y-\gamma)^2 1(X \le \theta) + (Y-\delta)^2 1(X > \theta)]\,.$$

- Population parameters at stage n:

$$(\gamma_n, \delta_n, \theta^n) = \arg\min_{\gamma,\delta,\theta} P_n[(Y-\gamma)^2 1(X \le \theta) + (Y-\delta)^2 1(X > \theta)]\,.$$

- Under mild conditions:

$$(\alpha_n|\theta^n - \theta_0|, \sqrt{\alpha}_n|\gamma_n - \gamma_0|, \sqrt{\alpha}_n|\delta_n - \delta_0|) = O(1)\,.$$

The three parameter problem

Under mild assumptions

- When $\alpha_n = o(n)$,

$$\{n^{1/3}\alpha_n^{2/3}|\hat{\theta}^n - \theta^n|, n^{1/3}\alpha_n^{1/6}|\hat{\gamma}_n - \gamma_n|, n^{1/3}\alpha_n^{1/6}|\hat{\delta}_n - \delta_n|\} = O_P(1)\,.$$

- When $\alpha_n = n$,

$$\{n|\hat{\theta}^n - \theta^n|, \sqrt{n}|\hat{\gamma}_n - \gamma_n|, \sqrt{n}|\hat{\delta}_n - \delta_n|\} = O_P(1)\,.$$

- When $n = o(\alpha_n)$,

$$\{n|\hat{\theta}^n - \theta^n|, \sqrt{n}|\hat{\gamma}_n - \gamma_n|, \sqrt{n}|\hat{\delta}_n - \delta_n|\} = O_P(1)\,.$$