

Validation de modèles en statistique Bayésienne

Sophie Donnet ¹ et Adeline Samson ²

¹Céramade, Université Paris Dauphine, France

²Laboratoire MAP5, Université Paris Descartes, France

Applibugs, 4 juin 2009

Modèle

- ▶ x^{obs} données observées
- ▶ modèle paramétrique $\mathcal{M}_0 : X|\theta \sim L(x; \theta)$
- ▶ loi a priori $\theta \sim \pi(\theta)$

Question

- ▶ Observations x^{obs} compatibles avec \mathcal{M}_0 ?
- ▶ Observations x^{obs} surprenantes ou extrêmes pour \mathcal{M}_0 ?

Objectif

Outils de mesure de cette surprise

Introduction

Prédictive
p-values

Definition
Propriétés
Graphiques

Exemple

Modèle
calcul des p-values
illustration
numérique

Conclusion

Références

Outil usuel : p-value

Construction d'une p-value

- $t(\cdot)$ statistique de test qui étudie un aspect du modèle
- $m(\cdot)$ loi à choisir
- $\mathbb{P}^{m(\cdot)}$ probabilité calculée sous la loi $m(X)$ ou $m(t(X))$, en supposant \mathcal{M}_0 vrai

p-value

$$p = \mathbb{P}^{m(\cdot)}[t(X) \geq t(x^{obs})]$$

Interprétation

- ▶ Probabilité d'observer des données plus "extrêmes" que x^{obs} si \mathcal{M}_0 vrai
- ▶ Mesure de la surprise
- ▶ Pas besoin de spécifier de modèle alternatif \mathcal{M}_1

Critères

- ▶ Test d'un aspect de \mathcal{M}_0 (t pas naturel \Leftrightarrow problème mal défini)
- ▶ Grandes valeurs de $t \Rightarrow$ faible compatibilité de x^{obs} avec \mathcal{M}_0 (ou forte surprise)

2 catégories

- ▶ t statistique de test
- ▶ t mesure de discrepancy (peut dépendre de certains paramètres du modèle)

Exemples de statistiques de test t

- ▶ $t(x_1, \dots, x_n) = \max(x_i)$
- ▶ $t(x_1, \dots, x_n) = \textit{median}(x_i)$
- ▶ $t(x_1, \dots, x_n) = \textit{mean}(x_i)$

Exemples de mesures de discrepancy

- ▶ Paramètres $\theta = (\psi, \eta)$
- ▶ Mesure de discrepancy $t(X, \eta)$
- ▶ $\mathbb{P}^{m(\cdot)}$ probabilité calculée sous la loi $m(X, \eta)$

$$p = \mathbb{P}^{m(\cdot)}[t(X, \eta) \geq t(x^{obs}, \eta)]$$

- ▶ χ^2 discrepancy

$$t(x, \eta) = \chi^2(x, \eta) = \sum_{i=1}^n \frac{(x_i - E[x_i|\eta])^2}{\text{Var}[x_i|\eta]}$$

- ▶ minimum discrepancy $t_{min}(x) = \min_{\eta} t(x, \eta)$
- ▶ average discrepancy $t_{avg}(x) = \int t(x, \eta)\pi(\eta|x^{obs})d\eta$

Choix de m

θ connu : $m(x; \theta) = L(x; \theta)$ et $p = \mathbb{P}^{L(\cdot; \theta)}[t(X) \geq t(x^{obs})]$

θ inconnu

► *Plug-in* : estimateur $\hat{\theta}$, $m(x; \hat{\theta}) = L(x; \hat{\theta})$ et

$$p_{plug} = \mathbb{P}^{L(\cdot; \hat{\theta})}[t(X) \geq t(x^{obs})]$$

→ simple mais ne prend pas en compte l'incertitude sur θ

► *Elimination de θ* : *similar p-value* : statistique u suffisante pour θ ($\Rightarrow L(x|u(x_{obs}), \theta) \perp \theta$)

$$p_{sim} = \mathbb{P}^{L(\cdot | u(x_{obs}), \theta)}[t(X) \geq t(x^{obs})]$$

→ conditionné par l'existence de U

Choix de m dans le cas général

Cadre général : θ inconnu et pas de statistique suffisante pour θ

Apport du cadre bayésien

- ▶ loi a priori sur $\theta \sim \pi(\theta) \Rightarrow$ intégration par rapport à θ
- ▶ **Predictive distribution et p-values**

Différents choix de m possibles

- ▶ Prior predictive distribution
- ▶ Posterior predictive distribution
- ▶ Partial predictive distribution
- ▶ Conditional predictive distribution

Prior predictive p -value [popularisée par Box (1980)]

Prior predictive distribution = marginal likelihood (loi de nouvelles données x avant de voir x^{obs})

$$m_{prior}(x) = \int L(x; \theta) \pi(\theta) d\theta$$

Prior predictive p-value

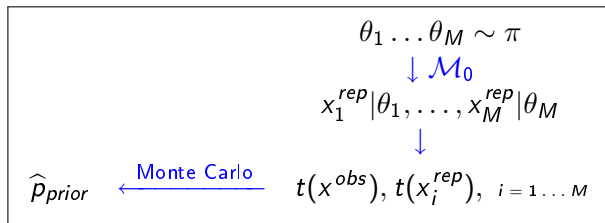
$$p_{prior} = \mathbb{P}^{m_{prior}(\cdot)}[t(X) \geq t(x^{obs})]$$

Avantage : Si $t(x) = \frac{1}{m_{prior}(x)}$, $p_{prior} = \mathbb{P}[m_{prior}(X) \leq m_{prior}(x^{obs})]$
 $m_{prior}(x^{obs})$ petit $\Rightarrow p_{prior}$ petit

Inconvénient : poids du prior important

- ▶ très bon modèle + "mauvais" prior $\Rightarrow p_{prior}$ petit
- ▶ si $\pi(\theta)$ impropre $\Rightarrow m_{prior}(x)$ impropre et calcul de p_{prior} impossible

Simulation sous $m_{prior}(x)$ par Monte-Carlo



Posterior predictive p -value [Guttman (1967), Rubin (1984)]

Validation de
modèles

Posterior predictive distribution (loi de nouv. données x après avoir vu x^{obs})

$$m_{post}(x|x^{obs}) = \int L(x; \theta)\pi(\theta|x^{obs})d\theta$$

Posterior predictive p -value

$$p_{post} = \mathbb{P}^{m_{post}(\cdot|x^{obs})}[t(X) \geq t(x^{obs})]$$

Avantages

- ▶ π propre ou impropre
- ▶ Plus influencée par le modèle que par le prior
si grand nombre d'observations, $m_{post} \approx \delta_{\hat{\theta}} \Rightarrow p_{post} \approx p_{plug}$.

Inconvénients : double utilisation des données pour obtenir $\pi(\cdot|x^{obs})$ et calculer p_{post}

Introduction

Prédictive
 p -values

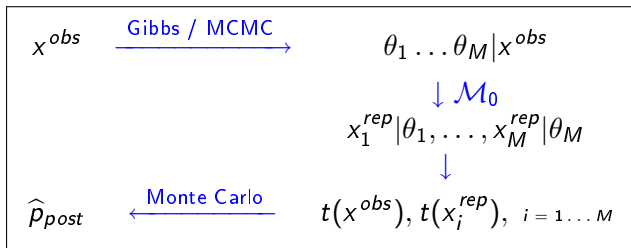
Définition
Propriétés
Graphiques

Exemple

Modèle
calcul des p -values
Illustration
numérique

Conclusion

Références



Extension de la posterior predictive p-value

[Meng (1994), Gelman, Meng, Stern (1996)]

- ▶ Si $\theta = (\psi, \eta)$ et validation de \mathcal{M}_0 en ψ , variable de discrepancy $t(x, \eta)$

$$m_{dpost}(x, \eta | x^{obs}) = \int L(x; (\psi, \eta)) \pi((\psi, \eta) | x^{obs}) d\psi$$

discrepancy posterior predictive p-value

$$p_{dpost} = \mathbb{P}^{m_{dpost}(\cdot | x^{obs})} [t(X, \eta) \geq t(x^{obs}, \eta)]$$

- ▶ $A(x)$ statistique auxiliaire constante (ex : taille de x)

$$m_{aux}(x | x^{obs}) = \int L(x | \theta, A(x) = A(x^{obs})) \pi(\theta | x^{obs}) d\theta$$

Partial posterior predictive p -value [Bayarri & Berger (2000)]

Partial posterior predictive distribution

- ▶ $t = t(x)$, $T = t(X)$
- ▶ $\pi(\theta|x^{obs} \setminus t(x^{obs})) \propto \pi(x^{obs}|t(x^{obs}), \theta)\pi(\theta)$

$$m(t|x^{obs} \setminus t(x^{obs})) = \int L(t|\theta)\pi(\theta|x^{obs} \setminus t(x^{obs}))d\theta$$

Partial posterior predictive p -value

$$p_{ppost} = \mathbb{P}^{m(t|x^{obs} \setminus t(x^{obs}))}(T \geq t(x^{obs}))$$

Avantages : pas de double usage des données car contribution de t^{obs} au calcul du posterior éliminée avant l'intégration contre θ

Validation de modèles

Introduction

Prédictive p -values

Définition

Propriétés

Graphiques

Exemple

Modèle

calcul des p -values

illustration

numérique

Conclusion

Références

Conditional predictive p-value

- ▶ u_1, u_2 statistiques représentant 2 aspects des données
- ▶ loi a posteriori pour θ à partir de $u_1(x^{obs})$

$$\pi(\theta|u_1(x^{obs})) \propto \pi(\theta)L(u_1(x^{obs}); \theta)$$

$$m_{cpred}(t|x^{obs}) = \int L(t|u_2(x^{obs}), \theta)\pi(\theta|u_1(x^{obs}))d\theta$$

Conditional posterior predictive p-value

$$p_{cpred} = \mathbb{P}^{m_p(t|x^{obs})}(T \geq t^{obs})$$

Choix de u_1 et u_2

- ▶ [Bayarri & Berger (2000)]

$$u_1(x^{obs}) = u_2(x^{obs}) = \hat{\theta}_{cMLE}(x) = \arg \max L(x|t; \theta)$$

- ▶ [Robert & Rousseau (2003)]

$$u_1(x^{obs}) = \hat{\theta}_{MLE}(x) \quad u_2(x^{obs}) = x^{obs} | \hat{\theta}_{MLE}(x)$$

Introduction

Prédictive
p-values

Definition

Propriétés

Graphiques

Exemple

Modèle

calcul des *p*-values

illustration

numérique

Conclusion

Références

Propriétés d'une *p*-value

Propriété d'une p-value

$p = p(x^{obs})$ vue comme la réalisation d'une variable aléatoire $p(X)$

Cadre fréquentiste

$$p(X) \sim \mathcal{U}[0, 1]$$

⇒ idée de la taille relative de $p(x^{obs})$

Cadre bayésien

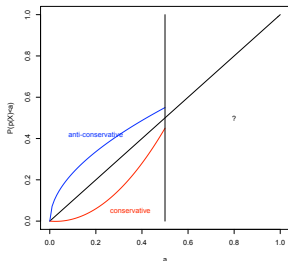
loi uniforme pas toujours vérifiée

Définitions des propriétés d'une p-value

Frequentist p-value : si $p(X) \sim \mathcal{U}[0, 1]$

Conservative si $\mathbb{P}(p(X) < \alpha) \leq \alpha \forall \alpha < 1/2$ sous \mathcal{M}_0

Anti-conservative si $\mathbb{P}(p(X) < \alpha) > \alpha \forall \alpha < 1/2$ sous \mathcal{M}_0



Asymptotiquement fréquentiste : si pour un échantillon

$$X_1, \dots, X_n, p(X) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{U}[0, 1] \forall \theta \in \Theta \text{ sous } \mathcal{M}_0$$

Propriétés des predictive p-values

[Bayarri, Berger, 2000, Robins, Van der Vaart, Ventura, 2000]

Propriétés de la prior predictive p-value

Soit $\pi(\theta)$ un prior propre. Sous \mathcal{M}_0 et sous la prior predictive distribution $m_{prior}(x)$, p_{prior} fréquentiste ($p_{prior} \sim \mathcal{U}[0, 1]$)

Propriétés de la posterior predictive p-value

Sous la prior predictive distribution $m_{prior}(x)$,

- ▶ p_{post} ne suit pas une loi $\mathcal{U}[0, 1]$
- ▶ p_{post} conservative : $\forall \alpha \leq \frac{1}{2}, \mathbb{P}(p_{post} \leq \alpha) \leq 2\alpha$
- ▶ p_{post} asymptotiquement fréquentiste

Propriétés des conditional predictive p-values

[Bayarri, Berger, 2000, Robins, Van der Vaart, Ventura, 2000]

Propriétés

- ▶ Soit $\pi(\theta)$ un prior propre. Sous certaines conditions, p_{cpred} fréquentiste
- ▶ Soit $\pi(\theta)$ un prior impropre. Sous certaines conditions, p_{cpred} fréquentiste
- ▶ p_{cpred} asymptotiquement fréquentiste

Introduction

Prédictive
 p -values

Définition
Propriétés

Graphiques

Exemple

Modèle
calcul des p -values
illustration
numérique

Conclusion

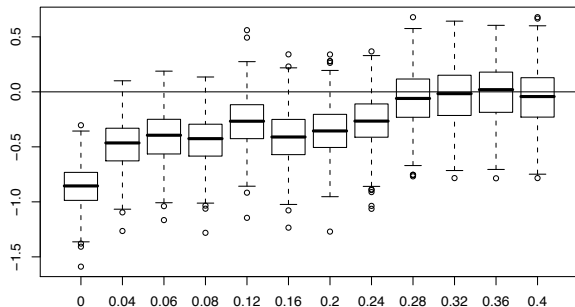
Références

Graphiques

Représentations graphiques [Gelman, 2004]

Détection de en quoi le modele est faux (contrairement à la p -value)

- ▶ Graphiques de réplifications des données
- ▶ Graphe de $D(x^{obs}, x^{rep}) = t(x^{obs}) - t(x^{rep})$



Introduction

Prédictive
 p -values

Definition
Propriétés
Graphiques

Exemple

Modèle
calcul des p -values
illustration
numérique

Conclusion

Références

Exemple sur un modèle hiérarchique

- ▶ [Bayarri & Castellanos(2007)]. *Bayesian Checking of the Second Levels of Hierarchical Models*. *Statistical Science* **Vol.22, no. 3, 363–367**
- ▶ [Gelman (2007)]. *Comment : Bayesian Checking of the Second Levels of Hierarchical Models*. *Statistical Science*. **Vol.22, no. 3, 349–352**

Modèle

Modèle

$$\begin{aligned}
 X_{ij}|\theta_i &\sim_{i.i.d.} \mathcal{N}(\theta_i, \sigma_i^2), \quad \forall i = 1 \dots l, \forall j = 1 \dots n_i \\
 \theta_i|\mu, \tau &\sim_{i.i.d.} \mathcal{N}(\mu, \tau^2) \\
 \pi(\mu, \tau^2) &= \pi(\mu)\pi(\tau^2) \propto \frac{1}{\tau}
 \end{aligned}$$

- ▶ Paramètres connus : σ_i^2
- ▶ Paramètres inconnus : θ_i, μ, τ^2

Problématique

- ▶ Adéquation du second niveau du modèle
 $\theta_i|\mu, \tau \sim \mathcal{N}(\mu, \tau^2)$
- ▶ Choix de t

$$t(X) = \max\{\bar{X}_1, \dots, \bar{X}_n\}$$

$$\text{avec } \bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$$

Introduction

Prédictive
p-valuesDefinition
Propriétés
Graphiques

Exemple

Modèlecalcul des p-values
illustration
numérique

Conclusion

Références

► Loi de $t(X)$

$$f_{t(X)}(t) = \sum_{i=1}^l N\left(t; \theta_i, \frac{\sigma_i^2}{n_i}\right) \prod_{l=1, l \neq i}^l F\left(t; \theta_k, \frac{\sigma_k^2}{n_k}\right)$$

N et F densités et f.d.r. de lois gaussiennes

► Vraisemblance intégrée

$$L(\mathbf{x}|\mu, \tau^2) = \prod_{i=1}^l N\left(\bar{x}_i; \mu, \frac{\sigma_i^2}{n_i} + \tau^2\right)$$
$$\Rightarrow \hat{\mu}, \hat{\tau}^2$$

Prior p-value

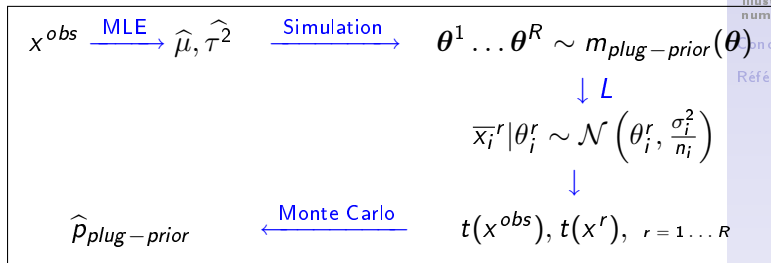


$$m_{prior}(\boldsymbol{\theta}) = \int L[\boldsymbol{\theta}|\mu, \tau^2] \underbrace{\pi(\mu)\pi(\tau^2)}_{\text{impropre}} d\mu d\tau^2$$

- ▶ Loi impropre donc pas possible de calculer p_{prior}

Prior plugin p-value

- ▶ $m_{plug-prior}(\theta) = m[\theta | (\mu, \tau^2) = (\hat{\mu}, \hat{\tau}^2)] = \prod_{i=1}^I N(\theta_i | \hat{\mu}, \hat{\tau}^2)$
- ▶ Pas d'expression explicite pour $p_{plug-prior}$
- ▶ En pratique :



Posterior predictive p-value

Simulation sous

$$m_{post}(x|x^{obs}) = \int L(x; \theta) \pi(\theta|x^{obs}) d\theta$$

où $\pi(\theta|x^{obs})$ = marginale de la posterior $\pi(\theta, \mu, \tau^2|x^{obs})$
obtenue par Gibbs

Algorithme

1. $\mu|\tau^2, \theta, x^{obs} \sim \mathcal{N}(E_\mu, V_\mu)$ avec $E_\mu = \frac{\sum_{i=1}^I \theta_i}{n_i}$ et $V_\mu = \frac{\tau^2}{I}$
2. $\tau^2|\mu, \theta, x^{obs} \sim \chi^{-2}(I-1, \tilde{\tau}^2)$ avec $\tilde{\tau}^2 = \frac{\sum_{i=1}^I (\theta_i - \mu)^2}{I-1}$
3. $\theta_i|\mu, \tau^2, x^{obs} \sim \mathcal{N}(E_i, V_i)$

Introduction

Prédictive
p-values

Definition
Propriétés
Graphiques

Exemple

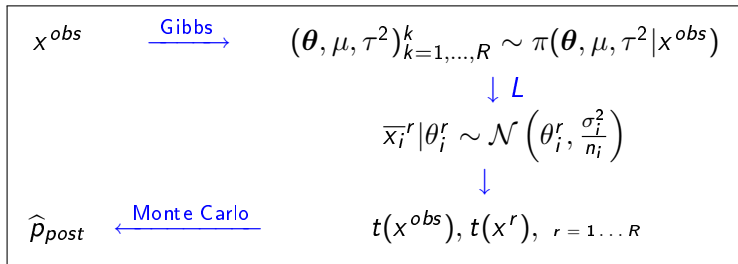
Modèle
calcul des p-values
illustration
numérique

Conclusion

Références

Posterior predictive p-value en pratique

En pratique :



Partial posterior predictive p-value

Simulation sous

$$m_{ppp}(x|x^{obs} \setminus t^{obs}) = \int L(x; \theta) \pi(\theta|x^{obs} \setminus t^{obs}) d\theta$$

où $\pi(\theta|x^{obs} \setminus t^{obs})$ marginale de la partial posterior

$$\pi(\theta, \mu, \tau^2|x^{obs} \setminus t^{obs}) \propto \frac{\pi(\theta, \mu, \tau^2|x^{obs})}{f(t^{obs}|\theta)}$$
 obtenue par Gibbs

Algorithme

1. $\pi(\mu|\tau^2, \theta, x^{obs} \setminus t^{obs}) \propto \pi(\mu|\tau^2, \theta, x^{obs})$ *gaussienne*
2. $\pi(\tau^2|\mu, \theta, x^{obs} \setminus t^{obs}) \propto \pi(\tau^2|\mu, \theta, x^{obs})$ *inverse χ^2*
3. $\theta_j|\mu, \tau^2, x^{obs} \setminus t^{obs} \propto \frac{\pi(\theta|\mu, \tau^2, x^{obs})}{f(t^{obs}|\theta)}$ *non simulable* \Rightarrow
Metropolis-Hastings

Partial posterior predictive p-value en pratique

En pratique :

$$X^{obs} \xrightarrow{\text{M-Hastings}} (\boldsymbol{\theta}, \mu, \tau^2)_{k=1, \dots, R} \sim \pi(\boldsymbol{\theta}, \mu, \tau^2 | X^{obs} \setminus t^{obs})$$

$\downarrow L$

$$\bar{x}_i^r | \theta_i^r \sim \mathcal{N}\left(\theta_i^r, \frac{\sigma_i^2}{n_i}\right)$$

\downarrow

$$t^{obs}, t^r, r = 1 \dots R$$

$$\hat{p}_{pppt} \xleftarrow{\text{Monte Carlo}}$$

Modèle de simulation

▶ $l = 5, n_j = 8.$

▶

$$X_{ij} \sim \mathcal{N}(\theta_i, 4), \forall i = 1 \dots 5, \forall j = 1 \dots 8$$

$$\theta_i \sim \Gamma(0.6, 0.2), \forall i = 1 \dots 5$$

Valeurs des p-values :

$$p_{plug-prior} = 0.121 \quad p_{post} = 0.381 \quad p_{ppp} = 0.011$$

Remarques :

- ▶ Seule p_{ppp} détecte le problème
- ▶ p_{post} très conservatrice, $p_{plug-prior}$ conservatrice

Introduction

Prédicative
p-values

Definition
Propriétés
Graphiques

Exemple

Modèle
calcul des p-values
**Illustration
numérique**

Conclusion

Références

Distribution sous \mathcal{M}_0

Validation de
modèles

Introduction

Prédicative
p-values

Définition
Propriétés
Graphiques

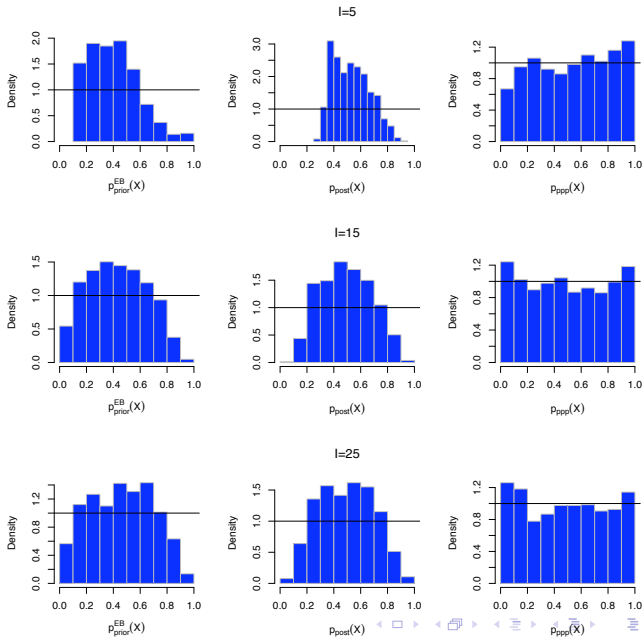
Exemple

Modèle
calcul des p-values

**Illustration
numérique**

Conclusion

Références



Conclusion

p-values

- ▶ Détecte à quel point les données sont extrêmes par rapport au modèle
- ▶ Grande variété (prior predictive, posterior predictive, conditional predictive, partial predictive)
- ▶ Qualités diverses
- ▶ Pas de consensus universel mais
 - ▶ p_{post} la plus facile à calculer et qualités "acceptables"
 - ▶ p_{ppp} la plus satisfaisante pour les qualités
- ▶ Extensions pour améliorer : Bootstrap p -value [Fraiser & Rousseau (2000)]

Graphiques

- ▶ Basés sur les réplifications des données
- ▶ Détecte en quoi le modèle est faux

Introduction

Prédictive
 p -values

Definition
Propriétés
Graphiques

Exemple

Modèle
calcul des p -values
Illustration
numérique

Conclusion

Références

Bibliographie

- ▶ **Meng** (1994), Annals of Stat, *Posterior predictive p-values*
- ▶ **Gelman, Meng, Stern** (1996), Stat Sinica, *Posterior predictive assessment of model fitness via realized discrepancies*
- ▶ **Bayarri, Berger** (1997), JSDS, *Measures of surprise in bayesian analysis*
- ▶ **Bayarri, Berger** (2000), JASA, *P values for composite null models*
- ▶ **Robins, Van der Vaart, Ventura** (2000), JASA, *Asymptotic distribution of p-values in composite null models*
- ▶ **Gelman** (2004), J Comput Graph Stat, *Exploratory data analysis for complex models*
- ▶ **Bayarri, Castellanos** (2007), Stat Science, *Bayesian checking of the Second Levels of Hierarchical Models*
- ▶ **Gelman** (2007), Stat Science, *Comment : Bayesian Checking of the Second Levels of Hierarchical Models.*