A Limit Result for the Prior Predictive

by

Michael Evans
Department of Statistics
University of Toronto

and

Gun Ho Jang
Department of Statistics
University of Toronto

Technical Report No. 1004 April 15, 2010

TECHNICAL REPORT SERIES

University of Toronto

Department of Statistics

A Limit Result for the Prior Predictive

Michael Evans and Gun Ho Jang
Department of Statistics
University of Toronto

Abstract

We establish results concerning the convergence of the prior predictive distribution. An application is made to the problem of checking for prior-data conflict.

Keywords: minimal sufficiency, prior predictive, convergence, prior-data conflict

1 Introduction

Suppose we have a model given by a collection of probability measures $\{P_\theta : \theta \in \Theta\}$ where $P_\theta(A) = \int_A f_\theta(x)\ \mu(dx)$, i.e., each P_θ is absolutely continuous with respect to a support measure μ on the sample space $\mathcal{X}$, with the density denoted by f_θ. With this formulation a prior Π leads to a prior predictive probability measure on $\mathcal{X}$ given by $M(A) = \int_\Theta P_\theta(A)\,\Pi(d\theta) = \int_A m(x)\,\mu(dx)$, where $m(x) = \int_\Theta f_\theta(x)\,\Pi(d\theta)$. If $T : \mathcal{X} \to \mathcal{T}$ is a minimal sufficient statistic for $\{P_\theta : \theta \in \Theta\}$, then it is well known that the posterior distribution for θ is the same whether we observe x or $T(x)$ and so we denote the posterior by $\Pi(\cdot\,|\,T)$. Furthermore, the conditional distribution of x given T is independent of θ and we denote this conditional measure by $P(\cdot\,|\,T)$. The joint distribution $P_\theta \times \Pi$ can then be factored as $P_\theta \times \Pi = M \times \Pi(\cdot\,|\,x) = P(\cdot\,|\,T) \times M_T \times \Pi(\cdot\,|\,T)$ where M_T is the marginal prior predictive distribution of T. If $f_{\theta T}$ denotes the marginal density of T, with respect to a support measure μ_T on $\mathcal{T}$, then $m_T(t) = \int_\Theta f_{\theta T}(t)\,\Pi(d\theta)$ denotes the density of M_T with respect to μ_T. If π denotes the density of Π, with respect to a support measure ν on Θ, then we can write $m_T(t) = \int_\Theta f_{\theta T}(t)\pi(\theta)\,\nu(d\theta)$.

Our concern here is with the behavior of M_T as the amount of data grows. A simple example illustrates the asymptotic behavior of this distribution that we might expect to hold in more general situations.

Example 1 (Location normal)**.** Suppose that $x = (x_1, \ldots, x_n)$ is a sample from a $N(\mu, 1)$ distribution where $\mu \in R^1$ is unknown. Then a minimal sufficient statistic is given by $T_n(x) = \bar{x}$ and $T_n(x)$ converges almost surely to the true value μ_* as $n \to \infty$. Suppose we put a $N(\mu_0, \sigma_0^2)$ prior on μ. The prior predictive distribution M_{T_n} is then easily obtained from $\bar{x} = \mu + \bar{z}$ where $\bar{z} \sim N(0, 1/n)$ independent of $\mu \sim N(\mu_0, \sigma_0^2)$, namely, M_{T_n} is the $N(\mu_0, \sigma_0^2 + 1/n)$ distribution. We see immediately that M_{T_n} converges in distribution to the $N(\mu_0, \sigma_0^2)$ distribution as $n \to \infty$. Furthermore, $m_{T_n}(t)$ converges almost surely to $(2\pi)^{-1/2}\sigma_0^{-1}\exp\{-(t-\mu_0)^2/2\sigma_0^2\}$ as $n \to \infty$, uniformly for t in a compact set. Simple computations show that these results do not depend on using a normal prior, namely, if we use prior measure Π with continuous density π, then M_{T_n} converges in distribution to Π as $n \to \infty$, and $m_{T_n}(t)$ converges almost surely to π as $n \to \infty$, uniformly for t in a compact set.

So in Example 1 we can think of $m_T(T(x))$ as a consistent estimator of the prior evaluated at the true value of the parameter. The significance of this is that the value of $m_T(T(x))$ gives an indication of whether or not the prior has been poorly chosen, in the sense that the true value of θ may lie in a region where little prior probability has been assigned. Of course, we cannot tell this from the value $m_T(T(x))$ itself but need to calibrate this on some scale. In Evans and Moshonov (2006, 2007) the P-value

$$M_T(m_T(t) \le m_T(T(x))), \tag{1}$$

and some variations of this, was proposed for checking for prior-data conflict. Note that this P-value is a modification of a P-value proposed by Box (1980) for general model checking in Bayesian contexts. In Example 1, when using the normal prior, we see that

(1) converges to $2(1-\Phi(|\mu_*-\mu_0|/\sigma_0))$ as $n\to\infty$, where μ_* is the true value of μ. So (1) is a consistent estimator of the P-value which measures whether the true value of the parameter lies in the tails of the prior. In Section 2 we prove that (1) converges to $\Pi(\pi(\theta)\le\pi(\theta_*))$, where θ_* is the true value of θ, in fairly general circumstances. The P-value $\Pi(\pi(\theta)\le\pi(\theta_*))$ will be small whenever the true value lies in a region of low prior probability and so we have an instance of prior-data conflict. As such, (1) is seen to be an appropriate measure of prior-data conflict.

A criticism of (1) is that, in the case of continuous models at least, the P-value is not invariant under smooth transformations. In particular, suppose that $W:\mathcal{T}\to\mathcal{W}$ is 1-1 and smooth. Let $J_W(t)$ be the reciprocal of the Jacobian determinant of W evaluated at t. Then, if instead of T we use $W(T)$ as the minimal sufficient statistic, the P-value is $M_W(m_W(w)\le m_W(W(T(x))))=M_T(m_T(t)J_W(t)\le m_T(T(x))J_W(T(x)))$ and this will not equal (1) unless $J_W(t)$ is constant.

In Evans and Jang (2010) the general problem of computing P-values, based on the density of a discrepancy statistic, to assess whether or not the data came from a single fixed distribution, was considered. An invariant P-value was proposed. For (1) this entails using instead the P-value

$$M_T(m_T^*(t)\le m_T^*(T(x))), \tag{2}$$

where $m_T^*(t)=m_T(t)E(J_T^{-1}(X)\,|\,T(X)=t), J_T(x)=|\det(dT(x)\circ dT'(x))|^{-1/2}$ and dT denotes the differential of T. The factor $E(J_T^{-1}(X)\,|\,T(X)=t)$ corrects for volume distortions due to the transformation T and is independent of θ because T is minimal sufficient. Note that $m_T^*(t)$ is the density of M_T with respect to the support measure $(E(J_T^{-1}(X)\,|\,T(X)=t))^{-1}\mu(dt)$. In Example 1 we have that $J_T(x)$ is constant and so (1) equals (2). While the P-value (2) will generally differ from (1), it is often the case that the effect of $E(J_T^{-1}(X)\,|\,T(X)=t)$ is negligible. We establish a convergence result for (2) in Section 3.

While there are numerous discussions concerning asymptotics for a posterior analysis, for example, Walker (1969), Heyde and Johnstone (1979), and Chen (1985), there seem to be almost no discussions concerned with convergence issues associated with the prior predictive distribution. Such results also have implications for methods that choose the prior based on the prior predictive. This paper addresses some of these problems.

2 Convergence of the Basic P-value

In the Appendix we provide the proof of the following result.

Theorem 1. Suppose Θ is an open subset of a Euclidean space and assume

(i) $T_n \to \theta$ a.s. P_θ for every θ,

(ii) $m_{T_n}(t) \to \pi(t)$ uniformly on compact subsets of Θ,

(iii) π is continuous and the prior distribution of $\pi(\theta)$ has no atoms,

then $M_{T_n}(m_{T_n}(t) \leq m_{T_n}(T_n(x_n))) \to \Pi(\pi(\theta) \leq \pi(\theta_*))$ a.s. P_{θ_*} where θ_* is the true value of θ.

Note that Theorem 1 implicitly assumes that the sampling model for T_n is continuous. We will subsequently discuss how to handle the discrete case.

To apply this result we need to establish (ii). We discuss several examples.

Example 2 (Scale-Gamma)**.** Let $x = (x_1, \ldots, x_n)$ be a sample from a Gamma(α_0, θ) distribution where the scale parameter $\theta > 0$ is unknown. Then the statistic $T_n(x) = (n\alpha_0)^{-1} \sum_{i=1}^n x_i \sim \text{Gamma}(n\alpha_0, \theta/(n\alpha_0))$ is minimal sufficient and it converges almost surely to the true value of θ. We prove the following result in the Appendix.

Lemma 2. If $T_n(x) \sim \text{Gamma}(n\alpha_0, \theta/(n\alpha_0))$ and the prior π on θ is continuous, then (ii) of Theorem 1 holds.

So if, in addition the prior distribution on $(0, \infty)$ has no atoms, then Theorem 1 applies and we have the convergence of (1). Certainly these conditions apply to the commonly used priors on a scale parameter.

The following example uses Example 2 in a problem of considerable importance for statistical practice.

Example 3 (Normal linear regression)**.** We consider first the situation where we have a sample $x = (x_1, \ldots, x_n)$ from a $N(\mu, \sigma^2)$ distribution with $\mu \in R^1$ and $\sigma > 0$ unknown. Then $T_n(x) = (T_{1n}(x), T_{2n}(x)) = (\bar{x}, s^2)$ is a minimal sufficient statistic and $T_n(x) \to (\mu, \sigma^2)$ as $n \to \infty$. We prove the following result in the Appendix.

Lemma 3. If $T_n(x) = (\bar{x}, s)$ where $\bar{x} \sim N(\mu, \sigma^2/n)$ independent of $s^2 \sim$Gamma$((n-1)/2, 2\sigma^2/(n-1))$ and the prior π on (μ, σ^2) is continuous, then (ii) of Theorem 1 holds.

The prior π is commonly prescribed by first stating a prior π_2 for σ^2 and then a conditional prior $\pi_1(\cdot \,|\, \sigma^2)$ for μ. As discussed in Evans and Moshonov (2006, 2007), it then makes sense to check π_2 first and, if π_2 passes, then check π_1. With this approach we can learn more about where the prior is deficient, if indeed there is a problem. Following that development, the check for π_2 is based on $T_{2n}(x) = s^2$ via the P-value $M_{T_{2n}}(m_{T_{2n}}(t_2) \le m_{T_{2n}}(T_{2n}(x)))$. Now $T_{2n}(x) \sim$ Gamma$((n-1)/2, 2\sigma^2/(n-1))$. So, when π_2 satisfies the conditions of Theorem 1, Example 2 applies with $\alpha_0 = 1/2$ and $M_{T_{2n}}(m_{T_{2n}}(t_2) \le m_{T_{2n}}(T_{2n}(x))) \to \Pi_2(\pi_2(\sigma^2) \le \pi_2(\sigma_*^2))$. To check π_1 the relevant P-value to use is

$$M_{T_{1n}}(m_{T_{1n}}(t_1 \,|\, T_{2n}(x) = s^2) \le m_{T_{1n}}(T_{1n}(x) \,|\, T_{2n}(x) = s^2) \,|\, T_{2n}(x) = s^2). \tag{3}$$

Now consider $m_{T_{1n}}(t_1 \,|\, T_{2n} = t_2) = m_{T_n}(t_1, t_2)/m_{T_{2n}}(t_2)$. Since $m_{T_n}(t_1, t_2) \to \pi(t_1, t_2)$ and $m_{T_{2n}}(t_2) \to \pi_2(t_2)$ uniformly on compact sets, then we have that $m_{T_{1n}}(t_1 \,|\, T_{2n} = t_2) \to \pi(t_1, t_2)/\pi_2(t_2) = \pi_1(t_1 \,|\, t_2)$ uniformly on compact sets. Furthermore, the measures $M_{T_{1n}}(\cdot \,|\, T_{2n} = s^2)$ converge in distribution to $\Pi_1(t_1 \,|\, \sigma_*^2), m_{T_{1n}}(t_1 \,|\, T_{2n} = s^2)$ converges almost surely $N(\mu_*, \sigma_*^2)$ to $\pi_1(t_1 \,|\, \sigma_*^2)$, and $m_{T_{1n}}(\bar{x} \,|\, T_{2n} = s^2)$ converges almost surely $N(\mu_*, \sigma_*^2)$ to $\pi_1(\mu_* \,|\, \sigma_*^2)$. This implies the convergence almost surely $N(\mu_*, \sigma_*^2)$ of (3) to $\Pi_1(\pi_1(\mu \,|\, \sigma_*^2) \le \pi_1(\mu_* \,|\, \sigma_*^2) \,|\, \sigma_*^2)$.

For a normal linear regression model $y_n = X_n\beta + e$, where $X_n \in R^{n \times k}$ and $e \sim N_n(0, \sigma^2 I)$, we have that $T_n(y) = (T_{1n}(y), T_{2n}(y)) = (b, s^2)$ where $b = (X_n' X_n)^{-1} X_n' y$ and $s^2 = (n-k)^{-1} ||y - Xb||^2$. Under suitable conditions on the X_n matrices we have that $T_n \to (\beta, \sigma^2)$ almost surely. The convergence results for this situation then proceed just as in as in the location-scale case.

The following result, with the proof provided in the Appendix, is sometimes useful in establishing condition (ii) in Theorem 1.

Theorem 4. Suppose Θ is an open subset of R^k, and for any $\epsilon > 0, K \subset \Theta$ compact there exist c_1, $c_2 > 0$ and $N > 0$ such that

$$f_{\theta,T_n}(t) \leq c_1 e^{-c_2 n} \text{ whenever } t \in K, ||t-\theta|| > \epsilon, \text{ and } n \geq N. \tag{4}$$

The following are equivalent:

(a) for any prior Π with continuous density π, then $m_{T_n}(t) \to \pi(t)$ uniformly for $t \in K$

(b) for any compact $K \subset \Theta$ and $\epsilon > 0$, then

$$\int f_{\theta,T_n}(t) I(||t-\theta|| < \epsilon)\, d\theta \to 1 \text{ as } n \to \infty, \text{ uniformly for } t \in K. \tag{5}$$

We now consider an example where the distribution of T_n is on a discrete subset of R^1. In such a case we can't expect condition (ii) of Theorem 1 to hold at values of t where $\pi(t) > 0$ but $m_{T_n}(t) = 0$ for all n. Suppose, however, that T_n has a lattice distribution with lattice step equal to h. Then for $kh < t \leq (k+1)h$ we define $m_{T_n}^{cont}(t) = m_{T_n}((k+1)h)/h$ and 0 otherwise and treat $m_{T_n}^{cont}(t)$ as a density with respect to length measure. Since $T_n(x)$ is always on the lattice we see immediately that

$$M_{T_n}^{cont}(m_{T_n}^{cont}(t) \leq m_{T_n}^{cont}(T_n(x))) = M_{T_n}(m_{T_n}(t) \leq m_{T_n}(T_n(x))).$$

We can then apply Theorem 1 to $m_{T_n}^{cont}$ and this proves the convergence of (1). Note that $m_{T_n}^{cont}(t) = \int_\Theta f_{\theta,T_n}^{cont}(t)\, \Pi(d\theta)$ where $f_{\theta,T_n}^{cont}(t) = f_{\theta,T_n}((k+1)h)/h$ when $kh < t \leq (k+1)h$.

Example 4 (Binomial)**.** Suppose that $x = (x_1, \ldots, x_n)$ is a sample from a Bernoulli(θ) distribution where $\theta \in (0,1)$ is unknown. Then $T_n(x) = \bar{x}$ is minimal sufficient and converges to θ. For $t \in \{0, 1/n, \ldots, 1\}$ then $f_{\theta,T_n}(t) = P_\theta(T_n = t) = \binom{n}{nt}\theta^{nt}(1-\theta)^{n(1-t)}$. In this case $\bar{x}$ has a discrete distribution on the lattice with step size equal to $1/n$. In the Appendix we prove the following result.

Lemma 5. If $nT_n(x) \sim$ Binomial(n, θ) and π is a continuous on $(0,1)$, then (ii) of Theorem 1 holds for $m_{T_n}^{cont}$.

Therefore, $m_{T_n}^{cont}(t)$ converges to $\pi(t)$ uniformly on each compact set and (1) converges provided the prior π satisfies the conditions of Theorem 1.

One interesting case where π does not satisfy the conditions of Theorem 1 arises when $\theta \sim$ Uniform$(0,1)$ as the prior distribution of $\pi(\theta)$ has all of its mass at 1. In this case, however, we have that $m_{T_n}^{cont}(t) \equiv n/(n+1) \to 1$ uniformly for all $t \in (0,1)$ and moreover $M_{T_n}(m_{T_n}(t) \leq m_{T_n}(T_n(x))) = 1 = \Pi(\pi(\theta) \leq \pi(\theta_*))$ and so the convergence result is obvious.

3 Convergence of the Invariant P-value

We now consider the convergence of (2). As noted, this P-value is invariant under smooth transformations and will agree with (1) whenever T is linear or the sampling model for T is discrete. This applies in Examples 1, 2, and 4 but not in Example 3.

Example 5 (Normal linear regression)**.** Consider the location-scale case. Clearly $T_{1n}(x) = \bar{x}$ is linear, and so the P-value for checking π_1 agrees with the invariant version. But $T_{2n}(x) = s^2$ is nonlinear and so the P-value (1) for checking π_2 is not the same as the invariant version. In this case $dT_{2n}(x) = (2/(n-1))(x_1 - \bar{x}, \ldots, x_n - \bar{x})$ giving $J_{T_{2n}}(x) = |\det(dT_{2n}(x) \circ dT'_{2n}(x))|^{-1/2} = (\sqrt{n-1}/2)s^{-1}$ and so $E(J_{T_{2n}}^{-1}(X) \,|\, T_n(X) = (\bar{x}, s^2)) = (2/\sqrt{n-1})s$. Therefore, the invariant P-value is equal to $M_{T_{2n}}(m_{T_{2n}}(t_2)t_2^{1/2} \leq m_{T_{2n}}(T_{2n}(x))(T_{2n}(x))^{1/2})$ and this converges almost surely to $\Pi_2(\pi_2(\sigma^2)\sigma \leq \pi_2(\sigma_*^2)\sigma_*)$ (see Theorem 6).

The proof of the following result is virtually identical to that of Theorem 1.

Theorem 6. Suppose Θ is an open subset of a Euclidean space and assume

(i) $T_n \to \theta$ a.s. P_θ for every θ,

(ii) $w_n(t) = E(J_{T_n}^{-1}(X) \mid T_n(X) = t)$ is continuous and $a_n w_n(t) \to w(t)$ for some sequence a_n,

(iii) $a_n m_{T_n}(t) w_n(t) \to \pi(t) w(t)$ uniformly on compact subsets of Θ,

(iv) π is continuous and the prior distribution of $\pi(\theta)w(\theta)$ has no atoms,

then $M_{T_n}(m_{T_n}^*(t) \leq m_{T_n}^*(T_n(x_n))) \to \Pi(\pi(\theta)w(\theta) \leq \pi(\theta_*)w(\theta_*))$ a.s. P_{θ_*} where θ_* is the true value of θ.

Note that $\pi(\theta)w(\theta)$ is the density of Π with respect to the support measure $(w(\theta))^{-1}\nu(d\theta)$. Also note that when θ is k-dimensional and $\sqrt{n}(T_n - \theta)$ is asymptotically normal, then, in many cases, we can take $a_n = n^{k/2}$.

The developments in this paper have required that the minimal sufficient statistic T_n be a consistent estimator of θ. The existence of such a minimal sufficient statistic is guaranteed for exponential models. Suppose, however, that we reparameterize via the 1-1, smooth function Ψ, namely, $\psi = \Psi(\theta)$. Then we must replace T_n by $\Psi(T_n)$ for the convergence results to hold as stated. If Ψ is nonlinear, however, then (1) will typically depend on whether we use T_n or $\Psi(T_n)$, namely, it will implicitly depend on the parameterization. Using (2) this dependence is avoided and the P-value is independent of the choice of the minimal sufficient statistic or equivalently the parameterization. The use of (2) seems more appropriate than (1) for this reason, although there is typically very little difference in the P-values obtained.

4 Conclusions

We have established convergence results for various prior predictive P-values that show directly that these are appropriate for checking for prior-data conflict, namely, assessing if the true value of the parameter is in the tails of the prior. Essentially these results are restricted to situations where a version of the minimal sufficient statistic is a consistent estimate of the model parameter and this means our results apply in the context of exponential models. Similar results can undoubtedly be established in other contexts, in particular for group models, and these are currently being developed.

More generally convergence results for the prior predictive have implications for empirical Bayes methods. For example, suppose we have a family of priors $\{\pi_\alpha : \alpha \in \mathcal{A}\}$ with corresponding prior predictives m_{α,T_n} for a minimal sufficient statistic T_n. Then convergence of $m_{\alpha,T_n}(T_n(x))$ to $\pi_\alpha(\theta_*)$ has the implication that maximizing $m_{\alpha,T_n}(T_n(x))$ over α to select the prior is essentially finding the prior that has maximal value at the true value of the parameter. One might argue that it makes more sense to maximize $M_{\alpha,T_n}(m_{\alpha,T_n}(t) \le m_{\alpha,T_n}(T_n(x)))$ over α as then, based on the convergence of this P-value to $\Pi_\alpha(\pi_\alpha(\theta) \le \pi_\alpha(\theta_*))$, we are finding the prior for which the true value is least surprising. The implications of this are currently being investigated.

Appendix

Proof of Theorem 1

Let $\epsilon > 0$. Then there exists $\delta > 0$ such that $|\pi(t) - \pi(\theta_0)| < \epsilon/2$ whenever $t \in \bar{B}_\delta(\theta_0)$ and there exists N_1 such that for all $n > N_1, T_n(x_n) \in \bar{B}_\delta(\theta_0)$. Also there exists N_2 such that for all $n > N_2$ and for all $t \in \bar{B}_\delta(\theta_0)$, then $|m_{T_n}(t) - \pi(t)| < \epsilon/2$. So, if $n > \max\{N_1, N_2\}$ then $M_{T_n}(m_{T_n}(t) \le \pi(\theta_0) - \epsilon) \le M_{T_n}(m_{T_n}(t) \le m_{T_n}(T_n(x_n))) \le M_{T_n}(m_{T_n}(t) \le \pi(\theta_0) + \epsilon)$. Now we prove that $M_{T_n}(m_{T_n}(t) \le \pi(\theta_0) - \epsilon) \to \Pi(\pi(\theta) \le \pi(\theta_0) - \epsilon)$. Let $\epsilon' > 0$. Let $C \subset \Theta$ be compact such that $\theta_0 \in C$, $\Pi(\partial C) = 0$, and $\Pi(C) \ge 1 - \epsilon'/2$. By (i) and Slutsky's Theorem M_{T_n} converges in law to Π. Therefore, $M_{T_n}(C) \to \Pi(C)$ and so there exists N_3 such that for all $n > N_3, M_{T_n}(C) > 1 - \epsilon'$. Therefore, for all $n > N_3$, $M_{T_n}(m_{T_n}(t) \le \pi(\theta_0) - \epsilon) - \epsilon' \le M_{T_n}(m_{T_n}(t) \le \pi(\theta_0) - \epsilon, C) \le M_{T_n}(m_{T_n}(t) \le \pi(\theta_0) - \epsilon)$ and we can make the LHS and RHS as close as we like by choosing ϵ' small. Let $\epsilon'' > 0$. There exists N_4 such that for all $n > N_4$ then $|m_{T_n}(t) - \pi(t)| < \epsilon''$ for all $t \in C$. When $n > \max\{N_3, N_4\}$ then, $M_{T_n}(\pi(t) \le \pi(\theta_0) - \epsilon - \epsilon'', C) \le M_{T_n}(m_{T_n}(t) \le \pi(\theta_0) - \epsilon, C) \le M_{T_n}(\pi(t) \le \pi(\theta_0) - \epsilon + \epsilon'', C)$ and the LHS converges to $\Pi(\pi(t) \le \pi(\theta_0) - \epsilon - \epsilon'', C)$ while the RHS converges to $\Pi(\pi(t) \le \pi(\theta_0) - \epsilon + \epsilon'', C)$. By choosing ϵ'' small we can make these quantities as close to $\Pi(\pi(t) \le \pi(\theta_0) - \epsilon, C)$ as we like. This proves that $M_{T_n}(m_{T_n}(t) \leq \pi(\theta_0) - \epsilon, C) \to \Pi(\pi(t) \leq \pi(\theta_0) - \epsilon, C)$ and this establishes that $M_{T_n}(m_{T_n}(t) \leq \pi(\theta_0) - \epsilon) \to \Pi(\pi(t) \leq \pi(\theta_0) - \epsilon)$. A similar argument shows that $M_{T_n}(m_{T_n}(t) \leq \pi(\theta_0) + \epsilon) \to \Pi(\pi(t) \leq \pi(\theta_0) + \epsilon)$ and this completes the proof.

Proof of Lemma 2

Suppose $K \subset \Theta$ is a compact set in $(0, \infty)$. Then, there are $0 < a < b < \infty$ such that $K \subset [a, b]$. Fix $\epsilon > 0$ satisfying $\epsilon < \min\{a/3, 1\}$. We prove (4) and (5), then apply Theorem 4. For $\theta - t > \epsilon$, $t/\theta \leq t/(t+\epsilon) \leq 1 - \epsilon/(b+\epsilon)$. Also for $\theta - t < -\epsilon$, $t/\theta \geq t/(t-\epsilon) \geq 1+\epsilon/(a-\epsilon)$. Hence, $|t/\theta - 1| > \eta = \epsilon/(b+\epsilon)$ for $|t-\theta| > \epsilon$ and $t \in K$. Note $\Gamma(n\alpha_0) \leq \sqrt{2\pi}(n\alpha_0)^{n\alpha_0 - 1/2} e^{-n\alpha_0}$. Since ue^{-u+1} has peak 1 at $u = 1$, $\alpha_1 = \sup_{u:|u-1|>\eta} ue^{-u+1} < 1$. Also there exists $N_1 > 1$ such that $n^{1/2}\alpha_1^{n\alpha_0/2} \leq 1$ for all $n \geq N_1$. Let $u = t/\theta$. Then for $|t - \theta| > \epsilon$, we get $|u - 1| > \eta$ and $f_{\theta,T_n}(t) = [(n\alpha_0)^{n\alpha_0} e^{-n\alpha_0}/\Gamma(n\alpha_0)t](ue^{-u+1})^{n\alpha_0} \leq a^{-1}(n\alpha_0/2\pi)^{1/2}\alpha_1^{n\alpha_0} \leq a^{-1}(\alpha_0/2\pi)^{1/2} e^{-n2^{-1}\alpha_0 \log(1/\alpha_1)}$. Hence, (4) holds.

For (5), let $I_0 = \int f_{\theta,T_n(t)} I(|t-\theta| < \epsilon)\, d\theta \leq \int_0^\infty [(n\alpha_0)^{n\alpha_0}/\Gamma(n\alpha_0)] u^{n\alpha_0 - 2}\, e^{-n\alpha_0 u}\, du = n\alpha_0/(n\alpha_0 - 1)$. Also for $t \in K$, $I(|t - \theta| < \epsilon) \geq I(|t/\theta - 1| < \eta_1) \geq I(|t/\theta - 1| < \eta_1 n^{-1/2} \log n)$ where $\eta_1 = \epsilon/(b + \epsilon)$. Then,

$$I_0 \geq \frac{(n\alpha_0)^{n\alpha_0} e^{-n\alpha_0}}{\Gamma(n\alpha_0)} \int u^{n\alpha_0 - 2} e^{-n\alpha_0(u-1)} I(|u - 1| < \eta_1)\, du.$$

For $|u - 1| < \eta_1 n^{-1/2} \log n$, a lower bound of the logarithm of the integrand is given by $\log(u^{n\alpha_0 - 2} e^{-n\alpha_0(u-1)}) = -n\alpha_0(u - 1) + (n\alpha_0 - 2)\log(1 - (1 - u)) \geq -n\alpha_0(n - 1) - (n\alpha_0 - 2)(1 - u + (1 - u)^2/2 + |1 - u|^3) \geq -n\alpha_0(1 - u)^2/2 - (2 + \alpha_0(\log n)^2)|1 - u|$. The change of variable $v = \sqrt{n\alpha_0}(u - 1)$ gives

$$I_0 \geq \frac{(n\alpha_0)^{n\alpha_0 - 1/2} e^{-n\alpha_0}}{\Gamma(n\alpha_0)} \int e^{-v^2/2 - (2+\alpha_0(\log n)^2)|v|/\sqrt{n\alpha_0}} I(|v| < \eta_1 \alpha_0^{1/2} \log n)\, dv.$$

By Stirling's formula $(n\alpha_0)^{n\alpha_0 - 1/2} e^{-n\alpha_0}/\Gamma(n\alpha_0) \to (2\pi)^{-1/2}$. The integral converges to $\int e^{-v^2/2}\, dv$ by the Lebesgue dominated convergence theorem. Hence, $I_0 \to 1$ as $n \to \infty$. Thus (5) holds and, by Theorem 4, Theorem 1 (ii) holds.

Proof of Lemma 3

We prove (4) and (5). The density $f_{\mu,\sigma^2}(\overline{x}, s^2)$ of $T_n = (\overline{X}, S^2)$ is given by

$$\frac{((n-1)/2)^{(n-1)/2} e^{-(n-1)/2}}{(2\pi/n)^{1/2}\Gamma((n-1)/2)(s^2)^{3/2}} \Big[\frac{s^2}{\sigma^2} \exp\Big(-\frac{n-1}{n}(\frac{s^2}{\sigma^2} - 1) - \frac{(\overline{x} - \mu)^2}{\sigma^2}\Big)\Big]^{n/2}.$$

Let I_1 be the first part and I_2 be the part inside the brackets. To prove (4), fix a compact set K. Let $a = \inf\{s^2 : (\overline{x}, s^2) \in K\} > 0$ and $b = \sup\{s^2 : (\overline{x}, s^2) \in K\} < \infty$. Also we consider $0 < \epsilon < \min(a/3, 1)$. Then, Stirling's formula gives $I_1/n \to 1/[\pi(2s^2)^{3/2}] \leq 1/[\pi(2a)^{3/2}]$. For $(\overline{x}, s^2) \in K$ and $||(\overline{x}, s^2) - (\mu, \sigma^2)|| > \epsilon$, we have $|\overline{x} - \mu| > \epsilon/2$ or $|s^2 - \sigma^2| > \epsilon/2$. Then there exists $\eta > 0$ such that $|s^2/\sigma^2 - 1| > \eta$. Since $ve^{-(v-1)(n-1)/n}$ is unimodal having peak $e^{-1/n}n/(n-1)$ at $v = n/(n-1)$, an upper bound of I_2 is obtained at $s^2/\sigma^2 = 1 + \eta$ or $s^2/\sigma^2 = 1 - \eta$ provided by $n > 1 + 1/\eta$. So $I_2 \leq \max((1 - \eta)\exp(\eta n/(n-1)), (1 + \eta)\exp(-\eta n/(n-1))) < 1$. If $|s^2/\sigma^2 - 1| \leq \eta$, then $|\overline{x} - \mu| > \epsilon/2$. Thus $(\overline{x} - \mu)^2/\sigma^2 = (\overline{x} - \mu)^2(s^2/\sigma^2)/s^2 \geq (\epsilon/2)^2(1 - \eta)/b$. So $I_2 \leq e^{-1/n}(n/(n-1))\exp(-\epsilon^2(1-\eta)/4b) < 1$ when $n > (1 + \exp(-\epsilon^2(1-\eta)/4b))^{-1}$. Hence (4) holds.

The integration range $||(\overline{x}, s^2) - (\mu, \sigma^2)|| < \epsilon$ contains $|\overline{x} - \mu| < \epsilon/2$ and $|s^2 - \sigma^2| < \epsilon/2$. Again, this region contains $|\overline{x} - \mu|/\sigma < \eta_2 = \epsilon/(\epsilon + 2b)$. Then,

$$
\begin{aligned}
I_3 &= \int\int f_{\mu,\sigma^2}(\overline{x}, s^2) I(||(\overline{x}, s^2) - (\mu, \sigma^2)|| < \epsilon)\, d\mu\, d\sigma^2 \\
&\geq \int\int (2\pi\sigma^2/n)^{-1/2} \exp(-(n/2\sigma^2)(\overline{x} - \mu)^2) I(|\overline{x} - \mu|/\sigma < \eta_2) \times \\
&\quad \frac{((n-1)/2)^{(n-1)/2} s^{(n-1)/2-1}}{\Gamma((n-1)/2)(\sigma^2)^{(n-1)/2}} e^{-(n-1)s^2/2\sigma^2} I(|s^2 - \sigma^2| < \epsilon/2)\, d\mu\, d\sigma^2
\end{aligned}
$$

Using $v = \sqrt{n}(\overline{x} - \mu)/\sigma$, this integral can be separated into two parts, let I_4 and I_5 be the two integrals. Then, $I_4 = \Phi(\sqrt{n}\eta_2) - \Phi(-\sqrt{n}\eta_2) \to 1$ as $n \to \infty$, and $I_5 \to 1$ as $n \to \infty$ by Lemma 2 for $\alpha_0 = 1/2$. So $I_3 \to 1$ and (5) holds. Finally, Theorem 1 (ii) holds by Theorem 4.

Proof of Theorem 4

Suppose (a) holds. If $\epsilon > 0$ and K is compact, then $\int f_{\theta,T_n}(t)\pi(\theta) I(||t - \theta|| \geq \epsilon)\, d\theta \leq \int c_1 e^{-c_2 n} I(||t - \theta|| \geq \epsilon)\, \Pi(d\theta) \leq c_1 e^{-c_2 n} \to 0$ uniformly in $t \in K$. So if (a) holds, then $|\int f_{\theta,T_n}(t)\pi(\theta) I(||t - \theta|| < \epsilon)\, d\theta - \pi(t)| \to 0$ uniformly for $t \in K$. The set $K_\epsilon = \{\theta : ||\theta - \theta'|| \leq \epsilon$ for some $\theta' \in K\}$ is also compact when ϵ is small enough. Note that, for given ϵ, the convergence in (5) follows whenever this convergence holds for a smaller value of ϵ. Let π be a continuous density that is constant and positive on K_ϵ, then (b) follows.

Now suppose (b) holds, Π has a continuous density π, K is a compact subset of Θ and $\epsilon > 0$. Then K_ϵ is also compact for ϵ is small enough. Since π is uniformly continuous on K_ϵ, there exists $\delta_0 > 0$ such that $|\pi(\theta_1) - \pi(\theta_2)| < \eta/4$ whenever $\theta_1, \theta_2 \in K_\epsilon$ and $||\theta_1 - \theta_2|| < \delta_0$. From (b), there exists $L_1 > 0$ such that $|\int f_{\theta,T_n}(t) I(||t-\theta|| < \delta)\, d\theta - 1| < \eta/(4 \sup_{t\in K} \pi(t))$ for all $n \geq L_1$ and $t \in K$ where $\delta = \min(\delta_0, \epsilon)$. Also, there exist c_1, $c_2 > 0$ and $L_2 > 0$ such that $f_{\theta,T_n}(t) \leq c_1 e^{-c_2 n}$ whenever $t \in K, ||t - \theta|| \geq \delta, n \geq L_2$. Therefore, there exists L_3 such that $\int f_{\theta,T_n}(t)\pi(\theta) I(||t-\theta|| \geq \delta)\, d\theta \leq c_1 e^{-c_2 n} \leq \eta/4$ for all $n \geq L_3$. Finally, for $n \geq L = \max(L_1, L_2, L_3)$ and $t \in K$, we have that $|m_{T_n}(t) - \pi(t)| \leq \int f_{\theta,T_n}(t)\pi(\theta) I(||t-\theta|| \geq \delta)\, d\theta + \int f_{\theta,T_n}(t)|\pi(\theta) - \pi(t)| I(||t-\theta|| < \delta)\, d\theta + \pi(t)|\int f_{\theta,T_n}(t) I(||t-\theta|| < \delta)\, d\theta - 1| \leq \eta/4 + (\eta/4)(1 + \eta/4) + \eta/4 < \eta$ and we see that (a) holds.

Proof of Lemma 5

Since K is compact, there is $a > 0$ such that $0 < a \leq t \leq 1 - a < 1$ for all $t \in K$. For $t \in K$, Stirling's formula implies that $\log \binom{n}{nt} = -2^{-1} \log 2\pi n t(1-t) - n(t \log t + (1-t)\log(1-t)) + r(n,t)$ where $r(n,t) < (12n)^{-1} - (12nt+1)^{-1} - (12n(1-t)+1)^{-1} < 1$. So we have that $f_{\theta,T_n}(t) = (2\pi t(1-t)/n)^{-1/2} \exp(r(n,t)) \exp(ng(\theta,t))$ where $g(\theta,t) = t\log(\theta/t) + (1-t)\log((1-\theta)/(1-t))$. Let $0 < \epsilon < a/2$. Note $g(\theta,t)$ has maximum value 0 at $\theta = t$. Therefore, since $g(\theta,t)$ is continuous when $|t-\theta| > \epsilon$, we have that $b = -\sup_{t,\theta:|t-\theta|>\epsilon} g(\theta,t) > 0$. Also there is N_1 so that $n^{1/2}e^{-bn/2} \leq 1$ for $n \geq N_1$. So, when $t \in K$, $|t-\theta| > \epsilon$ and $n \geq N_1$, $f_{\theta,T_n}(t) \leq (2\pi a(1-a))^{-1/2} e \exp(-nb/2)$. Hence (4) holds.

Let $I_0 = \int_0^1 f^{cont}_{\theta,T_n}(t) I(|t-\theta| < \epsilon)\, d\theta$. Then, $I_0 \leq \int_0^1 f^{cont}_{\theta,T_n}(t)\, d\theta = n\binom{n}{nt}\int_0^1 \theta^{nt}\ (1-\theta)^{n(1-t)}\, d\theta = n/(n+1)$. Thus, using (4), we get

$$I_0 = \frac{n}{n+1} - \int_0^1 f^{cont}_{\theta,T_n}(t) I(|t-\theta| > \epsilon)\, d\theta \geq \frac{n}{n+1} - c_1 e^{-c_2 n} \to 1.$$

Hence, (5) holds. So Theorem 1 (ii) holds by Theorem 4.

References

Box, G. E. P., 1980. Sampling and Bayes' inference in scientific modelling and robustness. J. Roy. Statist. Soc. Ser. A 143 (4), 383–430, with discussion.

Chen, C. F., 1985. On asymptotic normality of limiting density functions with Bayesian implications. J. Roy. Statist. Soc. Ser. B 47 (3), 540–546.

Evans, M., Jang, G. H., 2010. Invariant P-values for model checking. Ann. Statist. 38 (1), 512–525.

Evans, M., Moshonov, H., 2006. Checking for prior-data conflict. Bayesian Anal. 1 (4), 893–914.

Evans, M., Moshonov, H., 2007. Checking for prior-data conflict with hierarchically specified priors. In: Upadhyay, A., Singh, U., Dey, D. (Eds.), Bayesian Statistics and its Applications. Anamaya Publishers, New Delhi, pp. 145–159.

Heyde, C. C., Johnstone, I. M., 1979. On asymptotic posterior normality for stochastic processes. J. Roy. Statist. Soc. Ser. B 41 (2), 184–189.

Walker, A. M., 1969. On the asymptotic behaviour of posterior distributions. J. Roy. Statist. Soc. Ser. B 31, 80–88.